**The Dissertation Committee for Susan C Wang**
**certifies that this is the approved version of the following dissertation:**

# STUDIES OF BACTERIAL CATABOLIC ENZYMES: IMPLICATIONS FOR THE EVOLUTION OF ENZYMES AND METABOLIC PATHWAYS

**Committee:**

**Christian P. Whitman, Supervisor**

**Creed W. Abell**

**Kevin N. Dalby**

**David W. Hoffman**

**Hung-wen Liu**

# STUDIES OF BACTERIAL CATABOLIC ENZYMES: IMPLICATIONS FOR THE EVOLUTION OF ENZYMES AND METABOLIC PATHWAYS

by

**Susan C Wang, B.S.**

**Dissertation**

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

The University of Texas at Austin

August 2003

**Acknowledgments**

I would like to thank both The University of Texas at Austin and the American Foundation for Pharmaceutical Education for their generous funding over the past five years. This work would not have been possible without their financial assistance.

There are a tremendous number of people (and pets!) whom I have to thank for their endless support during my graduate school experience. First and foremost I would like to express my gratitude to my advisor, Dr. Christian P. Whitman. His guidance, encouragement, consideration, and seemingly infinite patience over the past five years have been unbelievably helpful and motivational. I can only hope that my future experiences in science with my employer(s) are as enjoyable and intellectually stimulating. I would also like to thank the members of my dissertation committee for their insight and understanding. I would like to thank Dr. Stacy L. Stamps-Deanda, Dr. Robert M. Czerwinski, and Dr. William H. Johnson, Jr., who initiated me into the laboratory, taught me all of the techniques, and were great friends and colleagues. Thanks also go to Dr. William F. Waas, who provided a listening ear. A very special and important thank you goes out to Mr. Carl E. Dick, who has had to deal with increased difficulty over the past few years from someone inherently difficult. Thanks to those others (too many to list) who have come and gone – if you ever read this, you know who you are. Thank you to Alan, Fudge Marble, Latte, Midnight, Mocha, Molly, and Richard, who have made stress less stressful. Finally, thank you to my parents and my sister for their continuing support and love.

**STUDIES OF BACTERIAL CATABOLIC ENZYMES: IMPLICATIONS FOR THE EVOLUTION OF ENZYMES AND METABOLIC PATHWAYS**

Publication No. _____

Susan C Wang, Ph.D.
The University of Texas at Austin, 2003

Supervisor: Christian P. Whitman

The origins of metabolic pathways and the evolution of the enzymes that comprise them have provoked intense debate and spawned a number of theories. The "patchwork" theory of Jensen, in which existing enzymes are combined to give new pathways, is one. Another emerging theme is that of "catalytic promiscuity," the ability of an enzyme to catalyze a low-level activity that differs from its physiological function. Such an activity can be amplified through mutation(s) to yield a more efficient enzyme. These ideas are often used to explain the origins of "superfamilies," which consist of enzymes that catalyze different reactions yet share sequence and/or structural homology.

The catechol *meta*-fission pathway, a plasmid-encoded degradation pathway for simple aromatic compounds, is rich in enzyme chemistry and replete with

structural and evolutionary diversity. 4-Oxalocrotonate tautomerase (4-OT), the best characterized enzyme in this pathway, is a member of the tautomerase superfamily. Two additional enzymes, YwhB, an enzyme of unknown function, and *trans*-3-chloroacrylic acid dehalogenase (CaaD) are also members. One defining feature of this superfamily is the conservation of an N-terminal proline which functions as a catalytic base. CaaD catalyzes dehalogenation via a hydration mechanism. Through site-directed mutagenesis, kinetic and pH rate analysis, and irreversible inhibition with 3-halopropiolates, Pro-1 of the β-subunit of CaaD was identified as a general acid catalyst instead of a general base catalyst, thus differentiating it from the rest of the superfamily. 4-OT and YwhB also catalyze dehalogenation, but at a low level. These results suggest that one or both enzymes may be the ancestor(s) of CaaD. Both enzymes are also inhibited by 3-halopropiolates, but by a different mechanism than that observed for CaaD.

2-Hydroxymuconate semialdehyde dehydrogenase (2-HMSD) immediately precedes 4-OT in the *meta*-fission pathway and is a member of the aldehyde dehydrogenase superfamily. Its catalytic mechanism may utilize a cysteine to form a thiohemiacetal intermediate, which is followed by hydride transfer to $NAD^+$. 2-HMSD has been characterized with several substrates, and site-directed mutagenesis has identified the essential catalytic cysteine. Additionally, 2-HMSD is reversibly inhibited by a product analog. This work sets the stage for further investigaton of its structure and mechanism.

**Table of Contents**

# INTRODUCTION

## Retro-evolution of Metabolic Pathways.

The origin of metabolic pathways and the evolution of enzymes have intrigued scientists over the past several decades. Several hypotheses have been advanced to address these questions. Two of the most popular theories were suggested quite early – the retro-evolution or retrograde theory by Horowitz, and the recruitment theory proposed by Jensen (1-3). In 1945, Horowitz suggested that enzymes evolved in a backwards fashion. The first organisms might have lived in a "highly complex chemical environment" which provided all of the nutrients, such as vitamins, amino acids, and sugars, necessary for life (1). Over time, these nutrients became depleted. Organisms that were able to synthesize such compounds gained a selective advantage over those that could not. Initially, such syntheses were one-step biochemical pathways that allowed the organism to use an existing precursor to generate the necessary nutrient (Figure 1).

The expansion of pathways in this manner assumes a common link between proteins and the genes that encode them. Since the substrate of one reaction is the product of the immediately preceding reaction, one might suppose that the enzymes that catalyze both reactions share some active-site architecture. The "easiest" way to capitalize on such similarity is to duplicate an existing gene and modifiy it by mutation, creating a new protein to catalyze a preceding step. In Horowitz's view, Enzyme 2 (Figure 2) arises as a result of the duplication and mutation of Enzyme 1. These enzymes are homologous, meaning that they share sequence and/or structural similarity. Thus, as pathways grew in size and complexity, they expanded in a backwards or retrograde fashion. The "last" protein in the pathway (Enzyme 1, Figure 2) is the ancestral parent of all of the proteins that precede it in the series. Each additional new enzyme would confer an evolutionary advantage on the host

organism since it would provide another means for the host to synthesize needed materials.

Horowitz believed that his retrograde evolution theory gained more credibility when operons were discovered (2). An operon is defined as a series of genes, clustered together on a chromosome or a plasmid, that have a common control mechanism (4). Horowitz considered two possibilities for the origin of an operon. Either the genes of an operon may have originated in different regions of the chromosome(s) and were transposed, over time, into their positions in the final sequence, or they might have originated in series. Horowitz believed that the former hypothesis was unlikely. On the other hand, the latter hypothesis was both supported by and gave support to the retrograde theory. Since it was also shown at this time that the genes of a biosynthetic pathway were not required to be clustered in operons (5-7), Horowitz believed that the presence of an operon implied a common ancestor.

Horowitz's retrograde model is vulnerable in many aspects. Many intermediates in metabolism are so unstable that they would not exist long enough for an enzyme to evolve to use them as substrates (3). Another issue is the availablility of metabolic intermediates. As these compounds are removed from the environment, retrograde evolution becomes impossible (8). Moreover, if the retrograde theory were true in all cases, then every pathway would contain homologous enzymes. Although there are cases in which homologous enzymes are encoded sequentially in the genome, these are the exception and not the rule. One such example is found in the histidine biosynthetic pathway of bacteria, where two enzymes that share homology and catalyze successive reactions are consecutively encoded (9-11).

**(Potential) Retrograde Evolution: PFACRI and ImGPS**

In histidine biosynthesis, two enzymes involved in consecutive reactions are phosphoribosyl-formimino-5-aminoimidazole carboxamide ribonucleotide isomerase (PFACRI or HisA) and imidazole-3-glycerol phosphate synthase (ImGPS or HisF)

(Figure 3). These enzymes share 25% sequence identity and a common $(\beta/\alpha)_8$ barrel structural motif (9-11). Both enzymes share several conserved amino acids in their active sites and contain repeated structural patterns of symmetrical loops of the same size and conformation (12). A comparison of the N-terminal and C-terminal halves ("half-barrels") of each protein shows that invariant residues are located at structurally identical positions (11). These half-barrel structures are essentially superimposble, suggesting that both enzymes may have evolved from a common ancestral half-barrel (13, 14). ImGPS is able to catalyze the PFACRI reaction at a low level ($k_{cat}/K_m$ = 0.1 $mM^{-1}$ $s^{-1}$), but the reverse is not true (11). The structural, sequence, and mechanistic evidence indicates that this pair of enzymes support Horowitz's retrograde evolution theory.

**Catalytic Promiscuity and the Patchwork Evolution Model**

The ability of ImGPS to catalyze the PFACRI reaction at a low level is an example of what has been called "catalytic promiscuity" (15). This phenomenon, described in detail by O'Brien and Herschlag, involves the use of active-site features in a particular enzyme to catalyze an alternative and likely unintended reaction. Many enzymes exhibit catalytic promiscuity, which is one of the themes that will be addressed by the work described in this dissertation.

Such promiscuity lends credence to the second major theory regarding the evolution of metabolic pathways. First proposed by Ycas in 1974, the "patchwork evolution" model was refined by Jensen in 1976 (16, 3). Jensen suggested that early enzymes possessed very broad substrate specificity, allowing them to react with a wide range of substrates. This could have allowed early organisms to maximize their catalytic potential with the few enzymes they may have possessed. Over time, genes for these enzymes would be duplicated, permitting mutations to occur that would increase their specificity. The patchwork model assumes the existence of at least a

few early enzymes, the origins of which must have been from some other means such as that described in Horowitz's theories.

**Patchwork Evolution: The Pentachlorophenol Degradation Pathway**

The origins of a number of metabolic pathways can be explained by the patchwork model. One elegant example can be found in the soil bacterium *Sphingomonas chlorophenolica.* This particular organism is able to degrade pentachlorophenol (PCP), a xenobiotic pesticide, and use it as a sole source of carbon and energy (17). The degradation pathway is shown in Figure 4.

The pathway begins with the conversion of PCP to tetrachlorohydroquinone, which is catalyzed by PCP hydroxylase (PcpB) (18). Subsequently, TCHQ dehalogenase performs two reductive dehalogenations to form trichlorohydroquinone (TriCHQ) and then 2,6-dichlorohydroquinone (DCHQ) (19). The aromatic ring is then cleaved by an $Fe^{2+}$-dependent extradiol dioxygenase, PcpA (20).

The first enzyme in the pathway, PcpB, is a flavin monooxygenase that requires molecular oxygen and two equivalents of NADPH to convert PCP to TCHQ (18). This step of the pathway appears to be rate-limiting (17). Unlike most flavin monooxygenases, PcpB displays broad substrate specificity. It is thus likely that this enzyme was recently recruited to utilize PCP as a substrate and has not yet optimized its new function.

TCHQ dehalogenase is a member of the zeta class of the glutathione S-transferase (GST) superfamily (21, 22). GSTs typically catalyze the nucleophilic attack of glutathione to form a glutathione conjugate, which is often used by organisms to detoxify alkylating agents (23). One of the GSTs, maleylacetoacetate (MAA) isomerase, appears to be the precursor for TCHQ dehalogenase. Both enzymes share significant active site architecture, and TCHQ can catalyze the glutathione-dependent isomerization of a double bond in MAA with similar kinetic parameters to those measured for MAA isomerase (21, 24). These observations

suggest that an MAA isomerase was recruited to fulfill the dehalogenase role in this particular pathway. Indeed, an MAA isomerase has been discovered in *S. chlorophenolica* that appears to have the "promiscuous" ability to catalyze the TCHQ dehalogenation (24).

Copley has suggested that the original functions of PcpB and PcpA were to cleave naturally occurring chlorinated compounds such as 2,6-dichlorophenol (17). Chlorophenols are naturally produced by insects and fungi (25, 26), so it is reasonable for soil bacteria to develop the ability to degrade such compounds. A model for the patchwork assembly of the PCP degradation pathway is shown in Figure 5. PcpB, which accepts chlorophenols as substrate, hydroxylates PCP. Later, a MAA isomerase was recruited from an existing metabolic pathway to serve as the TCHQ dehalogenase. Finally, PcpA could perform its normal function as a ring cleavage dioxygenase.

**Recruitment of Existing Enzymes: Mandelate Racemase and Muconate-Lactonizing Enzyme**

The explanation for the origins of certain metabolic pathways using the recruitment model, as illustrated in the PCP degradation pathway, has been appropriated by others based on one of two lines of evidence. The first line of evidence is that a particular enzyme is "catalytically promiscuous" (15). Second, homologous enzymes can be shown to catalyze different reactions, thus suggesting a common ancestor. One of the first instances in which the latter strategy might have been used is in the evolution of mandelate racemase (MR) and muconate-lactonizing enzyme (MLE), enzymes found in the mandelate and β-ketoadipate pathways, respectively (Figure 6) (27). These pathways are found in various *Pseudomonas* species which can use either isomer of mandelate as a sole source of carbon and energy. Mandelate is believed to be a breakdown product of lignin, a component of

plant cell walls. The mandelate pathway generates benzoate, which funnels into the β-ketoadipate pathway and is converted to acetyl-CoA and succinate.

The first enzyme of the mandelate pathway is mandelate racemase, which catalyzes the racemization of R- and S-mandelate (Figure 7). It is a homooctamer in solution and has an overall molecular weight of 320 kD. This enzyme requires a divalent metal ion for activity, preferring magnesium over manganese (28). The reaction proceeds by a so-called two base mechanism. Residues on opposite sides of the substrate (Lys 166 and His 297) act as acids or bases depending on the direction of the isomerization. Lys-166 removes the proton from the carbon adjacent to the carboxylate (designated the α-proton), while His-297 protonates the intermediate. The substrate is a bidentate ligand, possessing two points at which it can attach to the metal ion. Mandelate thus binds one carboxylate oxygen and the α-hydroxyl to the metal ion (12).

Muconate-lactonizing enzyme is found in the β-ketoadipate pathway (Figure 6). MLE, which is quite common among bacteria, catalyzes the interconversion of muconolactone with *cis, cis*-muconate (29). The observed physiological reaction is the conversion of *cis*, *cis*-muconate to muconolactone (Figure 7). Like MR, MLE is an octamer and requires a divalent ion for activity, with $Mn^{2+}$ preferred over $Mg^{2+}$. The reaction is initiated by the abstraction of the α-proton, analogous to the first step in the MR-catalyzed reaction. A lysine residue (either 169 or 273) is thought to act as the general base (12).

In 1990 it was discovered that the crystal structures of MR and muconate lactonizing enzyme (MLE) were nearly superimposable (30). Both enzymes belong to the class of TIM-barrel enzymes which obtain their name from triose phosphate isomerase (TIM), the first enzyme shown to contain this particular structure. The TIM-barrel fold is an eightfold repeat of a β-strand followed by an α-helix, designated $(\beta/\alpha)_8$ (31). MR and MLE also share 26% sequence identity, which provides additional evidence for a possible common ancestor (27). Despite structural similarities that occur to the level of the active site, neither enzyme has been shown to

catalyze the other's reaction, even at a low level. However, both mechanisms involve the abstraction of the α-proton, leading to a stabilized enolate intermediate (32). It is thus likely that nature used a precursor with a similar active site structure and evolved it to fulfill the different functions that MR and MLE serve. Perhaps MLE itself, which is more prevalent in bacterial species than is MR, was the ancestral enzyme recruited to perform the racemization. Another, more likely, possibility is that an unknown ancestor gave rise to both enzymes due to its inherent catalytic promiscuity (27).

**Enzyme Superfamilies**

The enzyme "superfamily" was coined by John Gerlt at the University of Illinois and Patricia Babbitt at the University of California-San Francisco, and others (33). The discovery that mandelate racemase and muconate-lactonizing enzyme are homologous (in sequence and structure), although they catalyze different reactions, essentially began the discussion of enzyme superfamilies. This finding suggested that enzymes related by divergent evolution could catalyze different overall reactions (12). Enzyme superfamilies, which are a major theme of the work in this dissertation, involve a number of terms which are briefly outlined below.

A superfamily can be described in terms of either structure or function. Gerlt and Babbitt have proposed a definition that encompasses both. Terms in this dissertation that refer to superfamilies will be defined in the context Gerlt and Babbitt have suggested (12). An enzyme family is a group of homologous enzymes that catalyze the same reaction, both mechanistically and in terms of substrate. Members of a family often have 30% or greater sequence identity. Frequently these enzymes are orthologs; that is, they are the same enzyme found in different species. One example of a family of enzymes is the set of all triose phosphate isomerases. The TIMs found in different organisms catalyze the same reaction using the same substrate. An enzyme superfamily consists of groups of homologous enzymes that

catalyze either 1) the same reaction mechanistically but with different substrates or 2) different reactions sharing a common mechanistic feature. The latter is made possible by conserved active site residues sharing the same function in all superfamily members. Such enzymes often have less than 20% sequence identity.

**The Enolase Superfamily**

As described in the previous section, mandelate racemase and muconate-lactonizing enzyme are members of a superfamily. MR and MLE are homologous, and they catalyze different reactions despite sharing active site elements. Several other members of the MR/MLE superfamily, designated the enolase superfamily, have now been identified. Each enzyme catalyzes a different reaction, but they all share the common ability to abstract the α-proton of a carboxylate. These enzymes include the title enzyme, enolase, which catalyzes the β-elimination of water from 2-phosphoglycerate to yield phosphoenolpyruvate (34), as well as a number of other enzymes whose reactions are illustrated in Figure 8. The reactions involved are chemically difficult due to the high p$K_a$s of the protons involved. However, the ability of these enzymes to stabilize the enolate ion lowers the p$K_a$, enabling the reaction to occur (Figure 9).

**The Crotonase Superfamily – Stabilization of Enolate Anions**

The members of the crotonase ("enoyl CoA hydratase") superfamily share the common feature of the use of an oxyanion hole to stabilize enolate anions that arise during the hydrolysis of peptide bonds (36-38). Crotonase catalyzes the addition of water to enoyl CoA esters, yielding 3-(*S*)-hydroxyl acyl CoA esters (39-41). The reaction is the second step in the fatty acid β-oxidation pathway, which converts fatty acids to acetyl-CoA for use in the Krebs cycle. No functional groups apart from those

comprising the oxyanion hole are conserved among these enzymes (12). Reactions catalyzed by this superfamily are illustrated in Figure 10.

Many of the members of the crotonase superfamily use coenzyme A (CoA) esters as substrates. The structure of the oxyanion hole includes two perpendicular β-sheets that are surrounded by α-helices (12). In the coenzyme A-utilizing enzymes, the oxyanion hole is located at the opening of the active site, and the acyl group of the substrate enters a cavity that contains the catalytic groups. Three of these enzymes, crotonase, 4-chlorobenzoyl CoA dehalogenase, and $\Delta^{3,5}$, $\Delta^{2,4}$-dienoyl CoA isomerase, have structures that are again nearly superimposable (12). 4-chlorobenzoyl dehalogenase catalyzes nucleophilic aromatic substitution via an arylated enzyme intermediate (42-46). This reaction is part of a pathway that allows soil bacteria to degrade halogenated aromatic compounds. $\Delta^{3,5}$, $\Delta^{2,4}$-dienoyl CoA isomerase catalyzes a 1,5-proton transfer reaction which is utilized in the degradation of unsaturated fatty acids (47). The structural commonalities among these enzymes suggest that divergent evolution may have yielded these varying activities. For instance, experiments have succeeded in converting 4-chlorobenzoyl CoA dehalogenase into a low-level crotonase (47), indicating that at least these two enzymes may have had a common ancestor.

**The Radical SAM Superfamily**

One of the newest and least characterized superfamilies is denoted the radical *S*-adenosylmethionine (SAM) protein superfamily (49, 50). This superfamily is based on a common mechanistic feature of radical formation upon reductive cleavage of SAM using an iron-sulfur cluster. Radical SAM proteins catalyze a number of interesting, chemically difficult reactions including methylation, isomerization, ring formation, and protein radical generation. The best characterized members of this superfamily include lysine 2, 3-aminomutase (LAM), lipoic acid synthase (LipA), and biotin synthase (BioB). These reactions are illustrated in Figure 11. LAM is part

of a metabolic pathway that allows anaerobic bacteria to utilize lysine as both a carbon and a nitrogen source, as well as being part of a biosynthetic pathway for antibiotics in *Streptomyces* (50-54). Lipoic acid, the product of LipA, is required for the activity of several enzymes in the glycine cleavage system and the decarboxylation of α-keto acids, which are important metabolic processes (54). Biotin, which arises from the cyclization and insertion of sulfur into dethiobiotin by BioB, is an essential vitamin.

Previous work did not suggest that these proteins were homologous despite the presence of a conserved cysteine motif (56). Sofia, *et al.*, used intensive sequence analysis to detect a low level of sequence similiarity between these and other proteins, suggesting the presence of a common ancestor and predicting a common structural fold (49). No crystal structures have been solved for any of these enzymes, although they have been the subjects of intense investigation (56-60). However, the determination of one crystal structure will likely shed light on interesting evolutionary and mechanistic features for many, if not all, members of the radical SAM superfamily.

**The Catechol *meta*-Fission Pathway**

The catechol *meta*-fission (MF) pathway (Figure 12) is a plasmid-encoded set of enzymes that allows organisms to use various aromatic compounds as sole sources of carbon and energy. The particular plasmid of interest in the Whitman laboratory is the TOL (pWW0) plasmid from *Pseudomonas putida* mt-2, a soil bacterium (60). Aromatic compounds such as toluene and xylene are funneled into the pathway after enzyme-catalyzed conversion to catechol or substituted catechols. Catechol-2,3-dioxygenase (C2,3O) opens the ring, forming 2-hydroxymuconate semialdehyde (2-HMS). At this point, the MF pathway diverges. Substituents on 2-HMS tend to determine the branch that is used for further processing. In the multi-enzyme branch, 2-hydroxymuconate semialdehyde dehydrogenase (2-HMSD) oxidizes 2-HMS to the

corresponding dienol acid, 2-hydroxymuconate. Subsequently, 4-oxalocrotonate tautomerase (4-OT) converts the dienol to the conjugated ketone. Decarboxylation of the conjugated ketone is catalyzed by 4-oxalocrotonate decarboxylase (4-OD), yielding 2-hydroxy-penta-2,4-dienoate (HPD) (62). In the single enzyme branch, 2-hydroxymuconate semialdehyde hydrolase (2-HMSH) directly hydrolyzes 2-HMS to form HPD. The branches converge after the formation of HPD. Vinylpyruvate hydratase (VPH) adds water to HPD, resulting in the formation of 2-oxo-4-hydroxypentanoate (KHP). Finally, KHP aldolase converts 2,4-HP to acetaldehyde and pyruvate, which are channeled into the Krebs cycle.

The *meta*-fission pathway is an ideal model for the study of enzyme evolution for several reasons. Since it is plasmid-encoded, it can be inferred that the MF pathway was constructed in response to environmental pressures. It is likely that the enzymes of the TOL plasmid were recruited from existing metabolic pathways, in accordance with Jensen's hypothesis (3). Evolution of the MF pathway may share similiarities with that of the evolutionary scheme proposed by Copley for pentachlorophenol degradation (17). However, the MF pathway is highly optimized for degradation, in contrast to the pathway used for pentachlorophenol. Due to the differences in sizes, structures (that have been solved), and substrate requirements of these enzymes and the inherent instability of some of the pathway intermediates, it is not likely that this pathway evolved in the manner espoused by Horowitz's retrograde evolution theory.

Furthermore, many of the enzymes of this pathway show sequence similarity to enzymes from other organisms. For some of these enzymes, the sequence similarity was originally only obvious for orthologs. For instance, when 4-OT was initially investigated, only other 4-OTs from other bacteria apparently shared homology with it (63). As the field of bioinformatics has grown and genomic sequencing projects have been completed, it has become apparent that the enzymes of the *meta*-fission pathway also belong to enzyme superfamilies. 2-HMSD and 4-OT, two of the major topics of this dissertation, are prime examples of this phenomenon.

4-OT is a member of the tautomerase superfamily, and 2-HMSD is a member of the aldehyde dehydrogenase superfamily. These enzymes, and their respective superfamilies, are discussed below.

**The Aldehyde Dehydrogenase (ALDH) Superfamily**

The aldehyde dehydrogenases (ALDH) comprise an ubiquitous superfamily of enzymes found in both prokaryotes and eukaryotes. These divergently related enzymes metabolize a vast array of endogenous and exogenous aldehyde compounds (63, 64). Aldehyde compounds can often cause many detrimental effects to an organism, including cytotoxicity, mutagenicity, and carcinogenicity (65). Thus, ALDHs serve an important detoxification role in many biological systems. They also sometimes serve additional functions in the cell, including metabolism, hormone binding, nuclear restoration, and stress response (67-71). ALDHs are ~500 amino acids in size and associate as homodimers or homotetramers. They require either $NAD^+$ or $NADP^+$ for activity. Most ALDHs exhibit a strong preference, in terms of $K_m$, for one over the other. The mechanism of catalysis is thought to involve the formation of a thiohemiacetal intermediate followed by hydride transfer to the nucleotide cofactor. Broadly speaking, aldehyde dehydrogenases can be divided into seven categories: semialdehyde dehydrogenases, non-specifc ALDHs, betaine dehydrogenases, non-phosphorylating glyceraldehyde 3-phosphate dehydrogenases, phenylacetaldehyde dehydrogenases, lactaldehyde dehydrogenases, and ALDH-like proteins (72).

A number of crystal structures for aldehyde dehydrogenases have been determined to date (73-78). Although the enzymes crystallized do not share a great deal of sequence identity (~20%), their structures show certain common features. Overall, the structural folds appear similar despite the differences in primary and quaternary structure among enzymes. Differences are more apparent in the active site features, particularly in how ALDHs bind either substrate (73) or the nucleotide

cofactor (74-76). The Rossmann fold, a nucleotide binding domain commonly found among enzymes that use $NAD(P)^+$ or FAD, is present but possesses a slightly different structure than that found in most other dehydrogenases (74). In a typical $NAD^+$-dependent dehydrogenase, the Rossman fold contains six β-strands and five α-helices (the β-α-β motif) (79), but in ALDHs, the Rossmann fold is made of five β-strands connected by four α-helices (74).

**2-Hydroxymuconate Semialdehyde Dehydrogenase (2-HMSD)**

2-Hydroxymuconate semialdehyde dehydrogenase (2-HMSD), which is a member of the semialdehyde ALDH family, is the first enzyme of the multi-enzyme branch in the *meta*-fission pathway. A sequence alignment of five ALDH sequences whose structures have been solved with 2-HMSD (generated using CLUSTAL-W) is illustrated in Figure 13. Although the crystal structure of 2-HMSD has not yet been elucidated, the sequence identity shown allows a number of predictions to be made about the enzyme.

Due to both sequence identity and crystallographic studies, a general mechanism has been proposed for aldehyde dehydrogenases (Figure 14). In this mechanism, an active site cysteine (Cys-287 in 2-HMSD) attacks the aldehyde group, forming a thiohemiacetal intermediate. Other residues conserved in the active site are an asparagine residue, thought to stablilize the thiohemiacetal through hydrogen bonding (80, 81), and one of two conserved glutamate residues believed to act as a general base that deprotonates the active site cysteine (80, 82). The active site residues, denoted by (*), are depicted in Figure 15. This effect is thought to be mediated via a water molecule or a conformational change since structural work suggests that the glutamates are too far from the cysteine to directly deprotonate (82). Subsequently, hydride transfer to the nucleotide cofactor occurs, followed by release of the oxidized acid product. Despite the presence of the conserved glutamate residues in the *Vibrio harveyi* aldehyde dehydrogenase, His-450 (denoted by ‡) is

involved with the deprotonation (82). This was determined through site-directed mutagenesis of His-450 coupled with structural determination. This particular residue is not found in 2-HMSD, nor is it conserved among any of the other ALDHs whose structures have been solved. However, an active site histidine has also been implicated in the mechanism of aspartate β-semialdehyde dehydrogenase (78).

In addition to evolutionary questions, our interest in 2-HMSD involves a number of mechanistic questions. We were interested in determining whether the product of 2-HMSD is 2-hydroxymuconate. 2-HMS can exist as three isomers, the enol, the β, γ-unsaturated ketone, or its α, β-isomer. Assuming the three isomeras are present in aqueous solution, it would be interesting to determine which isomer is processed by the dehydrogenase, and which isomer is formed. If 2-HMSD accepts the conjugated aldehyde, in which the aldehyde is more electrophilic, and produces the conjugated form of 2-HM, 4-OT serves no apparent function. 2-HMSD has been studied previously (84-86). Only the native enzyme was studied; the enzyme was not cloned for over-expression. Moreover, its activities with various substrates were assessed only by UV-visible spectroscopy and not by NMR spectroscopy, which would provide more information. Although potential active site residues were identified, no mutagenesis or inactivation experiments were performed. In our studies, we cloned and overexpressed 2-HMSD, verified its kinetic parameters for 2-HMSD, characterized the reaction by NMR, obtained further insight as to its substrate specificity, and investigated the role of the putative catalytic cysteine in the mechanism (Cys-287).

**4-OT, the 4-OT Family, and the Tautomerase Superfamily**

4-OT follows 2-HMSD in the multi-enzyme branch of the catechol *meta*-fission pathway (Figure 12). This enzyme has been a subject of research in the Whitman laboratory for the past fifteen years. 4-OT is a homohexamer (trimer of homodimers) of 6810 Da subunits. It catalyzes a 1,3-keto-enol tautomerization (2-

hydroxymuconate to the β, γ-unsaturated ketone) as well as a 1,5-keto-enol tautomerization (2-hydroxymuconate to the α,β-unsaturated ketone) (Figure 16). The overall reaction is the conversion of the β, γ-unsaturated ketone to its α, β-isomer via a dienol intermediate. 4-OT is an ideal system for the of study structure-function relationships in enzymes for two major reasons. First, it catalyzes a proton transfer without the assistance of a metal ion or a cofactor, making it a relatively simple enzyme to study. Second, its small size (62 amino acids) facilitates its study using chemical synthesis and NMR spectroscopy, which would be difficult, if not impossible, for much larger proteins.

The current mechanism for the 4-OT-catalyzed reaction is illustrated in Figure 17. Pro-1 was established as the general base, primarily as a result of affinity-labeling studies, crystallography, chemical synthesis, and NMR studies (86-93). Although the N-terminal proline would not normally be thought of as a general base at physiological pH because its p$K_a$ is approximately 9.4, direct NMR titration of uniformly labeled $^{15}$N-enzyme has shown that the p$K_a$ is ~6.4. The 1000-fold decrease in p$K_a$ thus allows Pro-1 to function as a base.

In Figure 17, a single prime (′) indicates that the residue is found in the neighboring monomer within the same dimer, relative to Pro-1. A double prime (″) indicates that the residue is found in an adjacent dimer relative to Pro-1. Thus, the arginine-11 from a neighboring monomer was shown to be important in both binding and catalysis through site-directed mutagenesis (94, 95). Changing Arg-11 to an alanine increased $K_m$ ~9-fold and decreased $k_{cat}$ ~90 fold. Arg-11 likely acts as an electron sink, drawing electron density towards C-5 of the substrate and facilitating protonation.

Arginine-39, from a neighboring dimer, is important both for catalytic and structural reasons. Mutation to an alanine or a glutamine affects catalysis, the p$K_a$ of the N-terminal proline, and the overall structure of the enzyme (94, 95). Small changes are observed for $K_m$, but larger decreases (125-fold and ~400-fold for R39A and R39Q, respectively) are observed for $k_{cat}$. Arg-39 is involved in hydrogen

bonding with both the C-1 carboxylate and the C-2 oxygen, thus facilitating deprotonation at C-3. Mutation of this residue probably eliminates one of these hydrogen bonds, making deprotonation less favorable. Additionally, a slight increase of the p$K_a$ of Pro-1 is observed in the R39Q mutant. NMR titration indicates that the p$K_a$ is increased to 7.1. In this mutant, a larger proportion of Pro-1 is protonated and cannot function as a base in catalysis at cellular pH. Moreover, the R39Q mutant shows structural disruption, possibly allowing water into the largely hydrophobic active site.

Another critical residue in 4-OT is Phe-50 (96). The F50A mutant shows large changes in $K_m$ and $k_{cat}$ (11-fold increase and 175-fold decrease, respectively). Moreover, the F50A mutant shows large structural damage, suggesting that the phenylalanine residue is crucial for maintaining the hydrophobicity of the active site. The dielectric constant of the wild-type active site is estimated to be 14.7 ± 0.7, while that of the F50A mutant is estimated to be 21.6 ± 2.6. The active site of F50A 4-OT is more accessible to water than that of the wild-type, which may decrease the interactions between the polar active site residues and the substrate. The phenyl ring of Phe-50 may also interact with the diene functional group of 2-HM.

The 4-OT family of enzymes is the best characterized family in the tautomerase superfamily. Three members of this family are major subjects of this dissertation. Overall, the tautomerase superfamily is comprised of three families of enzymes. The title enzymes of these families are MIF (macrophage migration inhibitory factor), CHMI (5-carboxy-2-hydroxymuconate isomerase), and 4-OT (97). These enzymes share little sequence identity (~7%) but share structural homology. The common structural unit among these proteins is a β-α-β motif. In CHMI and MIF, each monomer consists of two β-α-β motifs that resemble the 4-OT dimer. Although 4-OT possesses six active sites, MIF and CHMI each contain three – one per monomer. Despite their low sequence identity, each protein contains an N-terminal proline that functions as a catalytic base. The reactions catalyzed by these enzymes are diagrammed in Figure 18.

MIF exists as a trimer of ~14 kD subunits. This protein acts as part of the inflammatory response (98). It also catalyzes two enzymatic reactions; the tautomerization of phenylpyruvate between its enol and keto forms (Figure 18), and the tautomerization of D-dopachrome to 5,6-dihydroxyindole-2-carboxylic acid, an intermediate in the biosynthesis of melanin (99, 100). Although the enzymatic properties of MIF have been extensively studied, they have not been correlated with MIF's anti-inflammatory activities (101).

CHMI catalyzes an isomerization in the homoprotocatechuate (HPC) pathway of *Escherichia coli* C. This pathway (Figure 19) allows *E. coli* C to use aromatic amino acids as its sole sources of carbon and energy (102). The HPC pathway closely parallels the *meta*-fission pathway (Figure 13). Unlike the MF pathway, which is found on a plasmid, the HPC pathway is chromosomally encoded (103). Like MIF, CHMI exists as a homotrimer of ~14 kD amino acid subunits. CHMI converts 5-carboxymethyl-2-hydroxymuconate (CHM) to 5-carboxymethyl-2-oxo-3-hexene-1,6-dioate (Figure 18), analogous to the 4-OT-catalyzed reaction.

**Homologues of 4-OT: YwhB from *Bacillus subtilis* and CaaD from *Pseudomonas pavonceae* 170**

The fact that 4-OT is part of a plasmid-encoded pathway is suggestive of its evolutionary origin. 4-OT might have evolved from a chromosomal enzyme possessing tautomerase or isomerase activities. It may also have arisen from an enzyme that possessed one of these activities in a low-level form (63). When 4-OT was first cloned in 1992, the only known homologues were isozymes from other species (104). With the successful completion of many genome sequencing projects, however, many chromosomal homologues of 4-OT have been identified (104). These homologues, all of small size (61-79 amino acids), are ubiquitous among the eubacteria and archaea. Most of these proteins are found in organisms that are not

known to degrade aromatic compounds, suggesting that they serve some other function.

A multiple sequence alignment among the three members of the 4-OT family that are at the crux of the work in this dissertation is shown in Figure 20. The conserved (and likely active site) residues include Pro-1, Arg-11, and Phe-50.

YwhB from *Bacillus subtilis* has been identified only as a putative tautomerase (106). Its true biological role remains unknown. When it was first discovered in 1998, it was the most closely related homologue to 4-OT, sharing 36% sequence identity. In solution, it exists as a homohexamer of 61 amino acids per subunit. YwhB contains Pro-1, Arg-11, and a tyrosine in place of Phe-50. Arg-39 is not present, but a lysine (Lys-37) is nearby. Experiments in our laboratory have shown that YwhB functions as both a tautomerase and an isomerase, although not with the same catalytic efficiency as 4-OT (63). As is the case for the characterized members of the tautomerase superfamily, Pro-1 is believe to act as a general base. A crystal structure of YwhB has been obtained, demonstrating that the active sites of YwhB and 4-OT are nearly superimposable (Figure 21) (63). The structural and sequence homology shared by these enzymes suggests a common evolutionary link, much like that shown earlier between mandelate racemase and muconate-lactonizing enzyme.

*trans*-3-Chloroacrylic acid dehalogenase (CaaD) from *Pseudomonas pavonceae* 170 catalyzes the conversion of *trans*-3-chloroacrylic acid to malonate semialdehyde via the addition of water (Figure 22). It also processes *trans*-3-bromoacrylic acid in the same fashion. This enzyme is part of a chromosomally-encoded pathway that allows *P. pavonceae* and other soil bacteria to utilize 1, 3-dichloropropene as a sole source of carbon (107-109). The degradation pathway is illustrated in Figure 23.

1, 3-dichloropropene is the active ingredient in the commercial fumigants Telone II, Telone C17, and Telone C35 (110). These xenobiotic compounds are often used to treat soil to kill nematode parasites prior to planting. 1, 3-dichloropropene

has only been in use for a few decades, so bacteria that process this compound likely evolved to do so very recently. CaaD is made up of two subunits (denoted the α- and β-subunits), both of which show sequence identity with 4-OT and YwhB (111). CaaD is reportedly a heterohexamer (a trimer of heterodimers) in solution, rather than a homohexamer like 4-OT and YwhB (111). Both subunits of CaaD contain the conserved N-terminal proline. Site-directed mutagenesis experiments indicate that the β-Pro-1 is essential for activity, but the α-Pro-1 is not (111). Arg-11 and Phe-50, conserved from 4-OT, are present in the α-subunit of the enzyme. Like YwhB, however, Arg-39 is not present in either subunit. In the absence of a crystal structure, no definite statements can be made about the active site of CaaD.

The sequence identity shown between these three enzymes raises some interesting evolutionary and mechanistic questions. What is the nature of the relationship between 4-OT, YwhB, and CaaD? It is reasonable to infer that all three enzymes shared a common ancestor. Since CaaD presumably evolved more recently, could one (or both) of the other enzymes have been recruited for the 1, 3-dichloropropene degradation pathway? If recruitment did occur, is it simply because the active-site structure is convenient for the task? Perhaps the active site structure was not designed to perform the dehalogenation reaction but evolved through a series of mutations to do so. Do any of the three enzymes exhibit catalytic promiscuity? Does YwhB or 4-OT show low levels of dehalogenase activity? Alternatively, can CaaD act as a tautomerase or an isomerase? Does Pro-1 of the β-subunit of CaaD function as a general base as in the other members of the tautomerase superfamily? These questions will be addressed in the following chapters.

The studies that follow have both functional and intellectual significance. The characterization of 2-HMSD will expand our understanding of aldehyde dehydrogenases in general, and the category of semialdehyde dehydrogenases in particular. The only present structure of a semialdehyde dehydrogenase is that of aspartate β-semialdehyde dehydrogenase (78). This particular enzyme is much smaller (367 amino acids) than and does not align well with 2-HMSD (486 amino

acids). Additionally, aspartate β-semialdehyde dehydrogenase binds NADP/H, not NAD/H. The work described here thus sets the stage for the determination of a crystal structure for 2-HMSD, which would be the first $NAD^+$-dependent semialdehyde dehydrogenase structure available. Although studies thus far have indicated that conserved cysteine, glutamate, and asparagine residues are "required" for catalysis, the elucidation of more recent structures has implicated unconserved residues such as histidine in catalysis. Further investigation of 2-HMSD and other ALDHs may identify other residues, conserved or not, that are important in catalysis and substrate specificity. Since ALDHs serve essential detoxification roles, among others in cells, such scrutiny is clinically and practically important. The capabilities of these enzymes may also be exploited for bioremediation purposes. Although there has been hesitation about releasing genetically modified organisms into the environment, it might be possible to engineer, for instance, *Pseudomonas putida* to overexpress the *meta*-fission pathway. The engineered bacteria – which does not express any foreign proteins – might then be used to detoxify locations contaminated with toluene or other aromatic compounds.

Studies of the relationships and mechanisms of 4-OT, YwhB, and CaaD contribute to the understanding of enzyme evolution. With the escalating abundance of antibiotic-resistant bacteria, the way in which enzymes evolve merits intensive study. It is clear that over a fairly brief period of exposure to a particular compound, bacterial enzymes and even entire pathways will evolve to utilize it if beneficial or avoid it if detrimental. The hardiness of bacteria, coupled with their high mutation, recombination, and growth rates, allows them to adapt much more quickly than other organisms with the exception of viruses. The determination of factors that influence enzyme evolution and/or an indication of the manner in which enzymes evolve may enable such events to be predicted, optimized, and, in some cases, prevented.

**REFERENCES**


1. Horowitz, N. H. *Proc. Natl. Acad. Sci. USA,* 1945, **31**, 153-157.
2. Horowitz, N. H., in Evolving Genes and Proteins. 1965, 15-23.
3. Jensen, R. A. *Annu. Rev. Microbiol.*, 1976, **30**, 409-425.
4. Jacob, F. and Monod, J. *Cold Spring Harbor Symp. Quant. Biol.,* 1961, **26**, 193.
5. Gorini, L., Gundersen, W., and Burger, M. *Cold Spring Harbor Symp. Quant. Biol.,* 1961, **26**, 173.
6. Maas, W. K. *Cold Spring Harbor Symp. Quant. Biol.,* 1961, **26**, 183.
7. Vogel, H. J. *Cold Spring Harbor Symp. Quant. Biol.,* 1961, **26**, 163.
8. Lazcano, A. and Miller, S. L. *J. Mol. Evol.,* 1999, **49**, 424-431.
9. Fani, R., Lio, P., and Lazcano, A. *J. Mol. Evol.*, 1995, **41**, 760-774.
10. Fani, R., *et al. Gene,* 1997, **197,** 9-17.
11. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. *Science*, 2000, **289**, 1546-1550.
12. Gerlt, J. A. and Babbitt, P. C. *Annu. Rev. Biochem.,* 2001, **70**, 209-246.
13. Fani, R., Lio, P., Chiarelli, I., Bazzicalupo, M. *J. Mol. Evol.* 1994, 38: 489.
14. Thoma, R., Schwander, M., Liebl, W., Kirschner, K., Sterner, R. *Extremophiles,* 1998, **2,** 379.
15. O'Brien, P. J. and Herschlag, D. *Chem. Biol.*, 1999, **6**, R91-R105.
16. Ycas, M. *J. Theor. Biol.*, 1974, **44**, 145-160.
17. Copley, S. D. *Trends. Biochem. Sci.*, 2000, **25**, 261-265.
18. Xun, L., and Orser, C. S. *J. Bact.*, 1991, **173**, 4447-4453.
19. Xun, L., Topp, E., and Orser, C. S. *J. Bact.*, 1992, **174**, 8003-8007.
20. Xu, L., Resing, K., Lawson, S. L., Babbitt, P. C., and Copley, S. D. *Biochemistry*, 1999, **38**, 7659-7669.
21. McCarthy, D. L., Navarrete, S., Willett, W. S., Babbitt, P. C., and Copley, S. D. *Biochemistry*, 1996, **35**, 14634-14642.

22. Board, P. G., Baker, R. T., Chelvanayagam, G., and Jermiin, L. S. *Biochem. J.*, 1997, **328**, 929-935.
23. Armstrong, R. N. *Chem.. Res. Toxicol.*, 1997, **10**, 2-18.
24. Anandarajah, K., Kiefer, P. M., Jr., Donohoe, B. S., and Copley, S. D. *Biochemistry*, 2000, **39**, 5303-5311.
25. Gribble, G. W. *Environ. Sci. Technol.*, 1994, **28**, 311A-319A.
26. Teunissen, P. J., Swarts, H. J., and Field, J. A. *Appl. Microbiol. Biotechnol.*, 1997, **47**, 695-700.
27. Petsko, G. A., Kenyon, G. L., Gerlt, J. A., Ringe, D., and Kozarich, J. W. *Trends Biochem. Sci.*, 1993, **18**, 372-376.
28. Fee., J. A., Hegeman, G. D., and Kenyon, G. L. *Biochemistry*, 1974, **13**, 2528-2532.
29. Gaal, A. and Neujahr, H. Y. *J. Bact.*, 1979, **137**, 13-21.
30. Neidhart, D. J., Kenyon, G. L., Gerlt, J. A., and Petsko, G. A. *Nature*, 1990, **347**, 692-693.
31. Banner, D.W., *et al. Nature,* 1975, **255**, 609-14.
32. Babbitt, P. C., and Gerlt, J. A. *Adv. Prot. Chem.*, 2001, **55**, 1-28.
33. Babbitt, P. C., *et al. Science*, 1995, **267**, 1159-1161.
34. Babbitt, P. C., *et al. Biochemistry*, 1996, **35**, 16489-16501.
35. Holden, H. M., Benning, M. M., Haller, T., and Gerlt, J. A. *Acc. Chem. Res.*, 2001, **34**, 145-157.
36. Gerlt, J. A., and Babbitt, P. C. *Curr. Opin. Chem. Biol.*, 1998, **2**, 607-612.
37. Xiang, H., Luo, L. S., Taylor, K. L., and Dunaway-Mariano, D. *Biochemistry*, 1996, **35**, 8103-8109.
38. Babbitt, P. C. and Gerlt, J. A. *J. Biol. Chem.*, 1997, **272**, 30591-30594.
39. Yang, G., Liang, P. H., Dunaway-Mariano, D. *Biochemistry,* 1994, **33**, 8527-8531.
40. Taylor, K. L., *et al. Biochemistry*, 1995, **34**, 13881-13888.
41. Yang, G., *et al*. *Biochemistry*, 1996, **35**, 10879-10885.

42. Taylor, K. L., Xiang, H., Liu, R. Q., Yang, G., and Dunaway-Mariano, D. *Biochemistry*, 1997, **36**, 1349-1361.
43. Xiang, H., Dong, J., Carey, P. R., and Dunaway-Mariano, D. *Biochemistry*, 1999, **38**, 4207-4123.
44. Bahnson, B. J. and Anderson, V. E. *Biochemistry*, 1989, **28**, 4173-4181.
45. D'Ordine, R. L., Bahnson, B. J., Tonge, P. J., and Anderson, V. E. *Biochemistry*, 1994, **33**, 14733-14742.
46. Hofstein, H. A., Feng, Y., Anderson, V. E., and Tonge, P. J. *Biochemistry*, 1999, **38**, 9508-9516.
47. Luo, M. J., Smeland, T. E., Shoukry, K., and Schulz, H. *J. Biol. Chem.,* 1994, **269**, 2384-2388.
48. Xiang, H., Luo, L., Taylor, K. L., and Dunaway-Mariano, D. *Biochemistry*, 1999, **38**, 7638-7652.
49. Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F., and Miller, N. E. *Nucl. Acids. Res.* 2001, **29,** 1097-1106.
50. Magnusson, O.T., Reed, G. H., and Frey, P.A. *Biochemistry*, 2001, **40**, 7773-7782.
51. Costilow, R. N., Rochovansky, O. M., and Barker, H. A. *J. Biol. Chem.,* 1966, **241**, 1573-1578.
52. Chirpich, T. P., Zappia, V., Costilow, R. N., and Barker, H.A. *J. Biol. Chem.*, 1970, **245**, 1778-1789.
53. Stadtman, T. C. *Adv. Enzymol. Relat. Areas Mol. Biol.*, 1973, **38**, 413-448.
54. Charter, J. H., II, *et al*. *Biochemistry,* 1974, **13**, 1227-1233.
55. Reed, L. J., and Hackert, M. L. *J. Biol. Chem.*, 1990, **265**, 8971-8974.
56. Duin, E. C., *et al*. *Biochemistry*, 1997, **36**, 11811-11820.
57. Wu, W., *et al. Biochemistry*, 2000, **39**, 9561–9570.
58. Frey, P. A. and Moss, M. L. *Cold Spring Harbor Symp. Quant. Biol.*, 1987, **52**, 571–577.
59. Ollagnier-de Choudens, S., *et al. Biochemistry,* 2000, **39**, 4165–4173.

60. Miller, J. R., *et al. Biochemistry,* 2000, **39**, 15166–15178.

61. Horn, J. M., Harayama, S., and Timmis, K. N., *Mol. Microbiol*., 1991, **5**, 2459-2474.

62. Stanley, T. M. *et al*., *Biochemistry,* 2000, **39**, 718-726.

63. Whitman, C. P., *Arch. Biochem. Biophys.*, 2002, **402**, 1-13.

64. Lindahl, R. *Crit. Rev. Biochem. Mol. Biol.*, 1992, **27**, 283-335.

65. Vasiliou, V., Pappa, A., and Petersen, D. R. *Chem. Biol. Interact.*, 2000, **129**, 1-19.

66. Pereira, F., *et al. Biochem. Biophy. Res. Comm.*, 1991, **175**, 831-838.

67. Yamauchi, K., Nakajima, J., Hayashi, H., Horiuchi, R., and Tata, J. R. *J. Biol. Chem.*, 1999, **274**, 8460-8469.

68. Cenedella, R. J. *Ophthalmic Res.*, 2001, **33**, 210-216.

69. Liu, F., Cui, X., Horner, H. T., Weiner, H., and Schnable, P. S. *Plant Cell*, 2001, **13**, 1063-1078.

70. Guerrero, F. D., Jones, J. T., and Mullet, J. E. *Plant Mol. Biol.*, 1990, **15**, 11-26.

71. Stroeher, V. L., Boothe, J. G., and Good, A. G. *Plant Mol. Biol.*, 1995, **27**, 541-551.

72. Vasiliou, V., and Sophos, N. A. *Chem. Biol. Interact.*, 2003, **143-144**, 5-22.

73. Lamb, A. L. and Newcomer, M. E. *Biochemistry*, 1999, **38**, 6003-6011.

74. Liu, Z.-J., *et al. Nature Struc. Biol.*, 1997, **4**, 317-326.

75. Johansson, K., *et al. Prot. Sci.,* 1998, **7**, 2106-2117.

76. Ahvazi, B., *et al. Biochem. J.*, 2000, **349**, 853-861.

77. Rossmann, M. G., Moras, D., and Olsen, K. W. *Nature*, 1974, **250**, 194-199.

78. Blanco, J., Moore, R. A., Kabaleeswaran, V., and Viola, R. E., *Protein Science*, 2003, **12**, 27-33.

79. Hadfield, A. T., *et al. J. Mol Biol.*, 1999, **289**, 991-1002.

80. Hurley, T. D., Steinmetz, C. G., and Weiner, H. *Adv. Exp. Med. Biol.*, 1999, **463**, 15-25.

81. Hempel, J., *et al. Adv. Exp. Med Biol.*, 1999, **463**, 53-59.
82. Wang, X., and Weiner, H., *Biochemistry*, 1997, **34**, 237-243.
83. Zhang, L, Ahvazi, B., Szittner, R., Vrielink, A., and Meighen, E. *Biochemistry,* 2000, **39**, 14409-14418.
84. Shaw, J. P., and Harayama, S. *Eur. J. Biochem.*, 1990, **191**, 705-714.
85. Shaw, J. P., Schwager, F., and Harayama, S. *Biochem. J.*, 1992, **283**, 789-794.
86. Inoue, J., Shaw, J. P., Rekik, M., and Harayama, S. *J Bacteriol.*, 1995, **177**, 1196-201.
87. Whitman, C. P., *et al. JACS,* 1992, **114**, 10104-10110.
88. Fitzgerald, M. C., Chernushevich, I., Standing, K. G., Kent, S. B. H., and Whitman, C. P., *JACS*, 1995, **117**, 11075-11080.
89. Stivers, J. T., *et al*. *Biochemistry*, 1996, **35**, 803-813.
90. Johnson, W. H. Jr., Czerwinski, R. M., Fitzgerald, M. C., and Whitman, C. P. *Biochemistry*, 1997, **36**, 15724-15732.
91. Subramanya, H. S., *et al. Biochemistry*, 1996, **35**, 792-802.
92. Taylor, A. B., Czerwinski, R. M., Johnson, W. H. Jr., Whitman, C. P., and Hackert, M. L. *Biochemistry*, 1998, **37**, 14692-14700.
93. Fitzgerald, M. C., Chernushevich, I., Standing, K. G., Whitman, C. P., and Kent, S. B. H., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 6851-6856.
94. Harris, T. K., *et al. Biochemistry*, 1999, **38**, 12343-12357.
95. Czerwinski, R. M., *et al. Biochemistry*, 1999, **38**, 12358-12366.
96. Czerwinski, R. M., Harris, T. K., Massiah, M. A., Mildvan, A. S., and Whitman, C. P., *Biochemistry*, 2001, **40**, 1984-1995.
97. Murzin, A. G. *Curr. Opin. Struct. Biol.*, **6**, 386-394.
98. Bernhagen, J., Calandra, T., and Bucala, R., 1998, *J. Mol. Med.,* **76**, 151-61.
99. Rosengren, E., *et al. FEBS Lett.,*1997,**417**, 85-88.
100. Rosengren, E., *et al. Mol. Med.*, 1996, **2**, 143-149.
101. Hermanowski-Vosatka, A., *et al. Biochemistry*, 1999, **38**, 12841-12849.

102. Sparnins, F. L., Chapman, P. J., and Dagley, S., *J. Bact.*, 1974, **120**, 159-167.

103. Roper, D. I., Fawcett, T., and Cooper, R. A., *Mol. Gen. Genet.*, 1993, **23**, 241-250.

104. Chen, L. H., *et al. J. Biol. Chem.*, 1992, **267**, 17716-17721.

105. Almrud, J. J., *et al. Biochemistry*, 2002, **41**, 12010-12024.

106. Kunst, F., *et al.*, *Nature*, 1997, **390**, 249-256.

107. Hartmans, S., Jansen, M. W., van der Werf, M. J., and De Bont, J. A. M., *J. Gen. Microbiol.*, 1991, **137**, 2025-2032.

108. Van Hylckama Vlieg, J. E. T. and Janssen, D. B., *Biodegradation*, 1992, **2**, 139-150.

109. Poelarends, G. J., Wilkens, M., Larkin, M. J., van Elsas, J. D., and Janssen, D. B., *Appl. Environ. Microbiol.*, 1998, **64**, 2931-2936.

110. Ou, L.-T., Thomas, J. E., Chung, K.-Y., and Ogram, A. V., *Biodegradation*, 2001, **12**, 39-47.

111. Poelarends, G. J., Saunier, R., and Janssen, D. B., *J. Bact.*, 2001, **183**, 4269-4277.

**Figure 1**. Depiction of a single enzyme evolving via Horowitz's retrograde hypothesis. Enzyme 1 is able to convert Precursor 1 into the desired product.

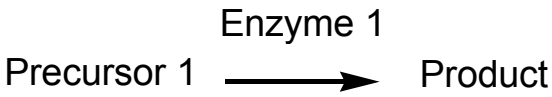


**Figure 2**. Depiction of two enzymes evolving via Horowitz's retrograde hypothesis. Enzyme 2 evolves from gene duplication of Enzyme 1 followed by mutation.

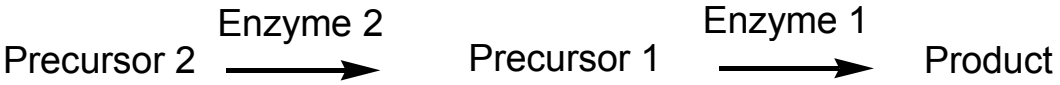


**Figure 3**. Catalysis by two enzymes involved in consecutive reactions of histidine biosynthesis. Top, phosphoribosyl-formimino-5-aminoimidazole carboxamide ribonucleotide isomerase (PFACRI; HisA). Bottom, imidazole-3-glycerol phosphate synthase (ImGPS; HisF).

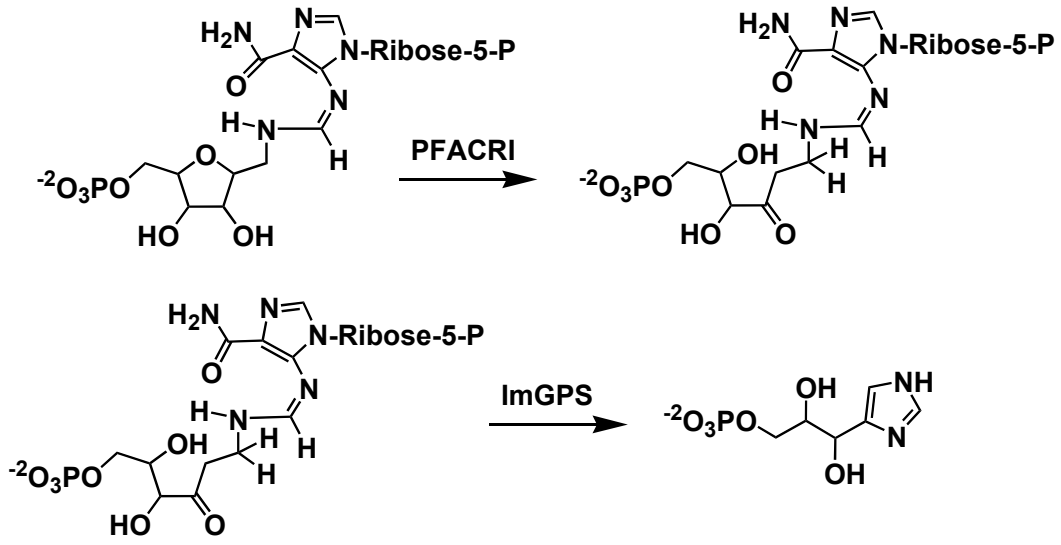


**Figure 4.** Pathway for degradation of pentachlorophenol. Abbreviations: PcpB, pentachlorophenol hydroxylase; TCHQ DH, trichlorohydroqunione dehalogenase; PcpA, dioxygenase.

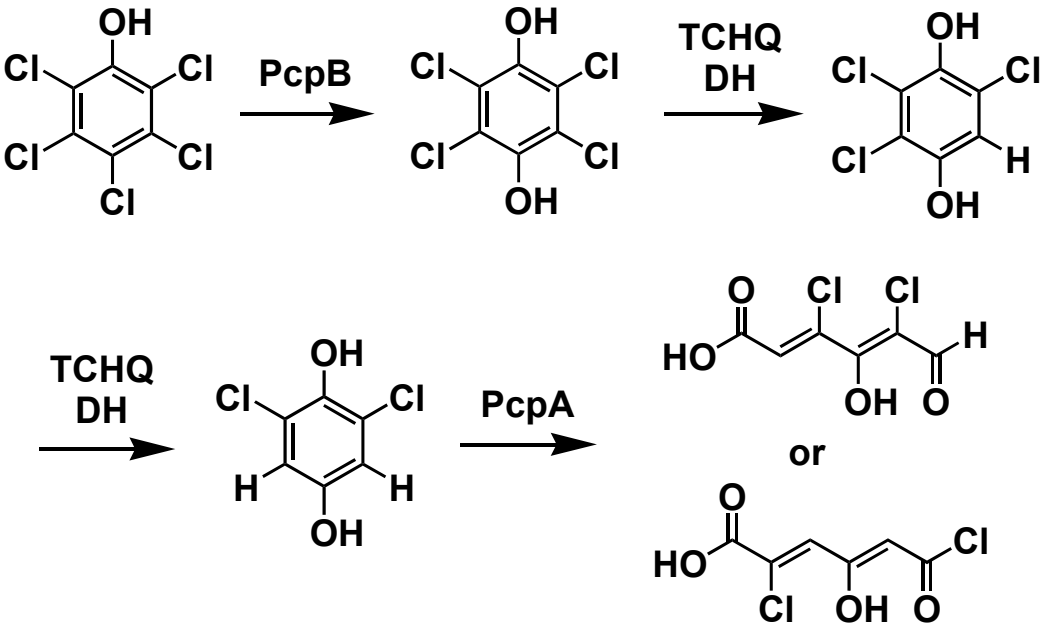


**Figure 5.** Proposed evolution of the pentachlorophenol degradation pathway. Abbreviations: PcpB, pentachlorophenol hydroxylase; TCHQ DH, tetrahydroquinone dehalogenase; MAAI, maleylacetoacetate isomerase; PcpA, dioxygenase (17).

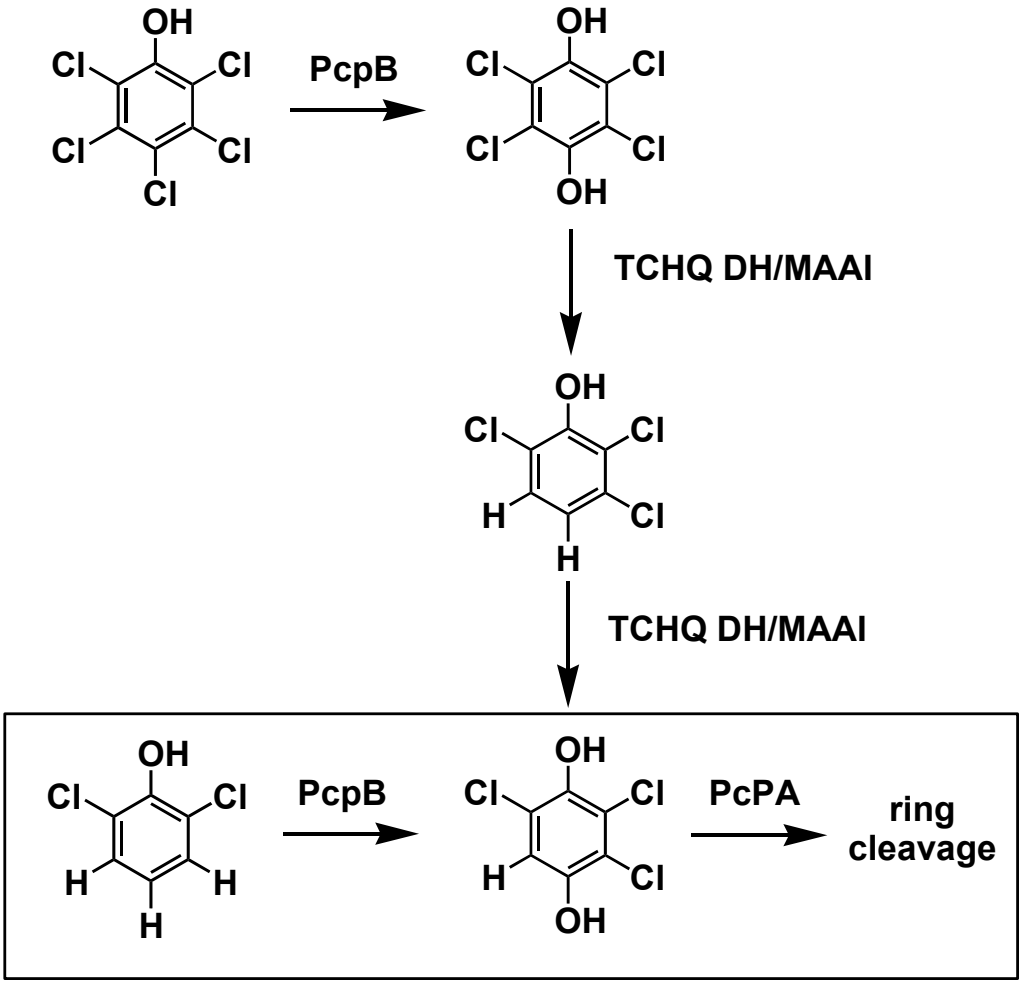


**Figure 6.** Top: pathway for mandelate degradation. Bottom: pathway for homoprotocatechuate degradation. Abbreviations: MR, mandelate racemase; MDH, *S*-mandelate dehydrogenase; BFD, benzoylformate decarboxylase; BADH, benzaldehyde dehydrogenase; BDO, benzoate dioxygenase; DH/DC, dehydrogenase/decarboxylase; C1,2O, catechol 1,2-dioxygenase; MLE, muconate-lactonizing enzyme; MLI, muconolactone isomerase; ELH, enol lactone hydrolase; CAT, coenzyme A transferase; SH-ase, thiolase

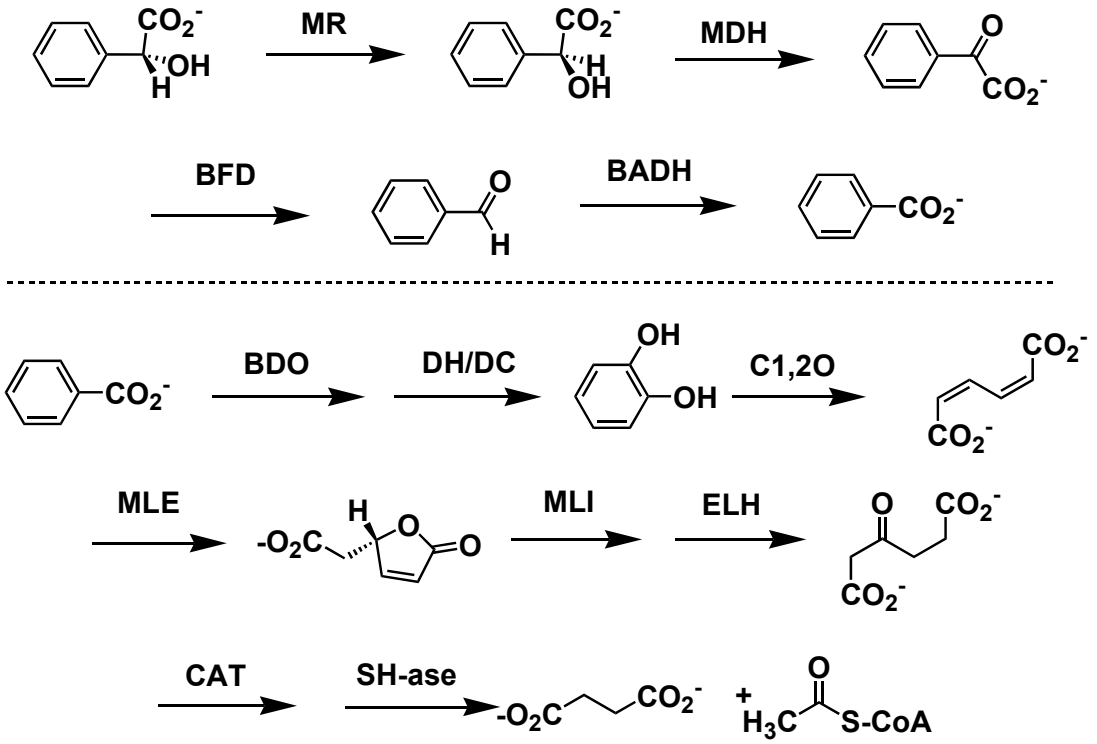


**Figure 7.** Top, reaction catalyzed by mandelate racemase (MR). Bottom, reaction catalyzed by muconate-lactonizing enzyme (MLE).

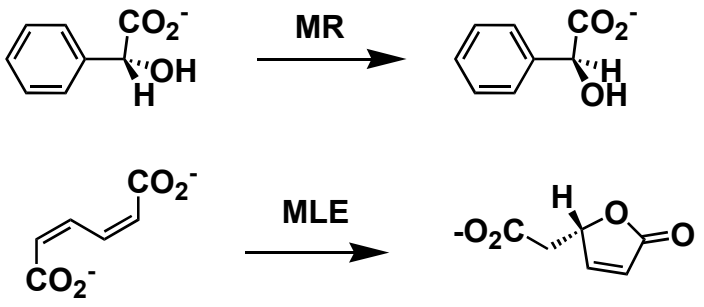


**Figure 8.** Reactions catalyzed by enolase superfamily members (32). Abbrevations: MR, mandelate racemase; MLE I, muconate-lactonizing enzyme; GalD, (D)-galactonate dehydratase; GlucD, (D)-glucarate dehydratase.

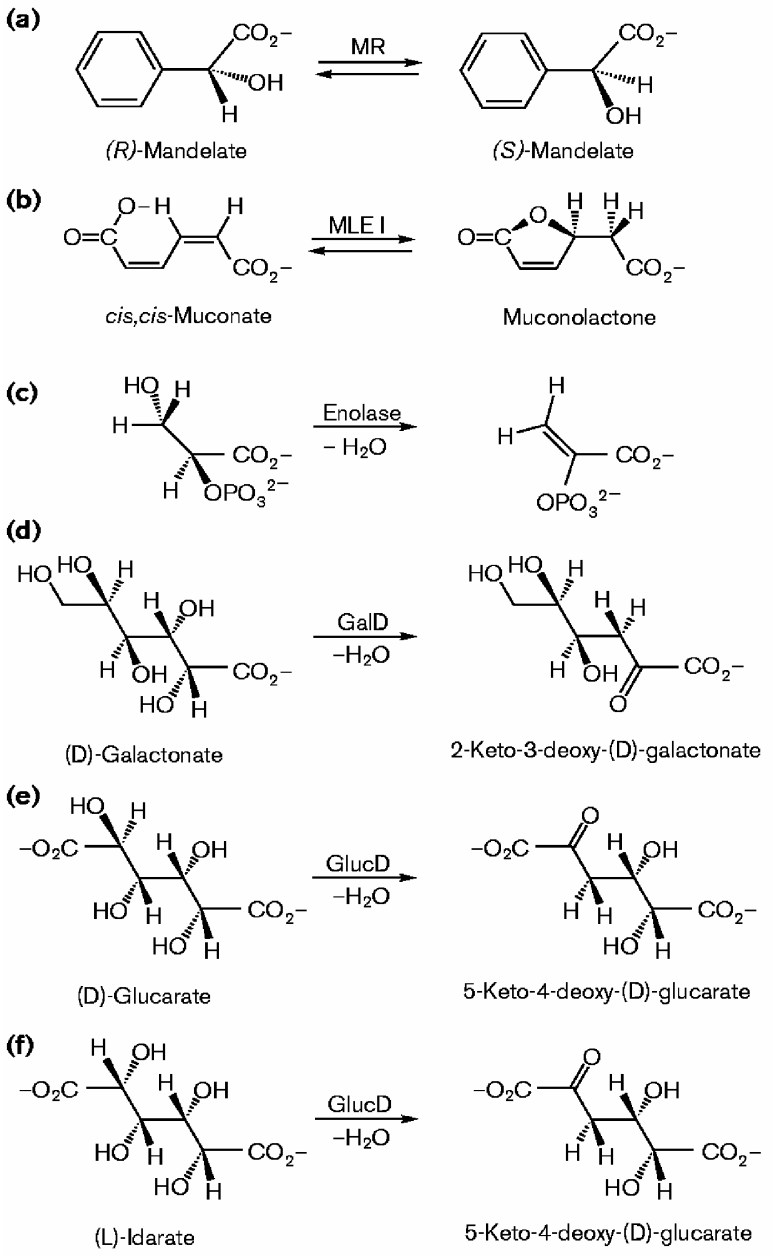


**Figure 9.** Proposed stabilization of the oxyanion hole by crotonase (35).

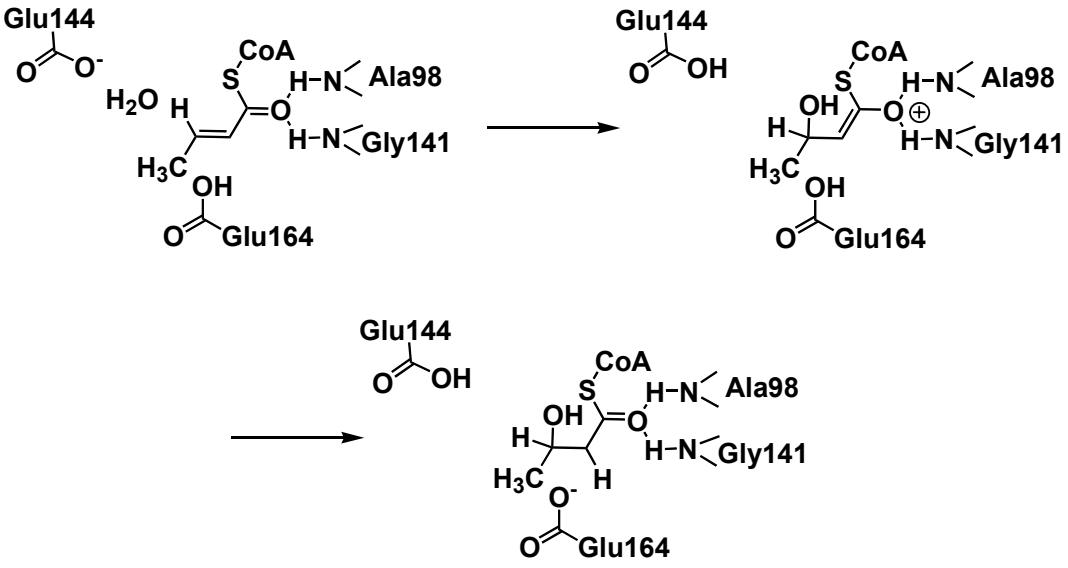


**Figure 10.** Reactions catalyzed by crotonase superfamily members (32). Abbreviations: DHNA, dihydroxynaphthoate; CaiD, carnitinyl CoA epimerase.

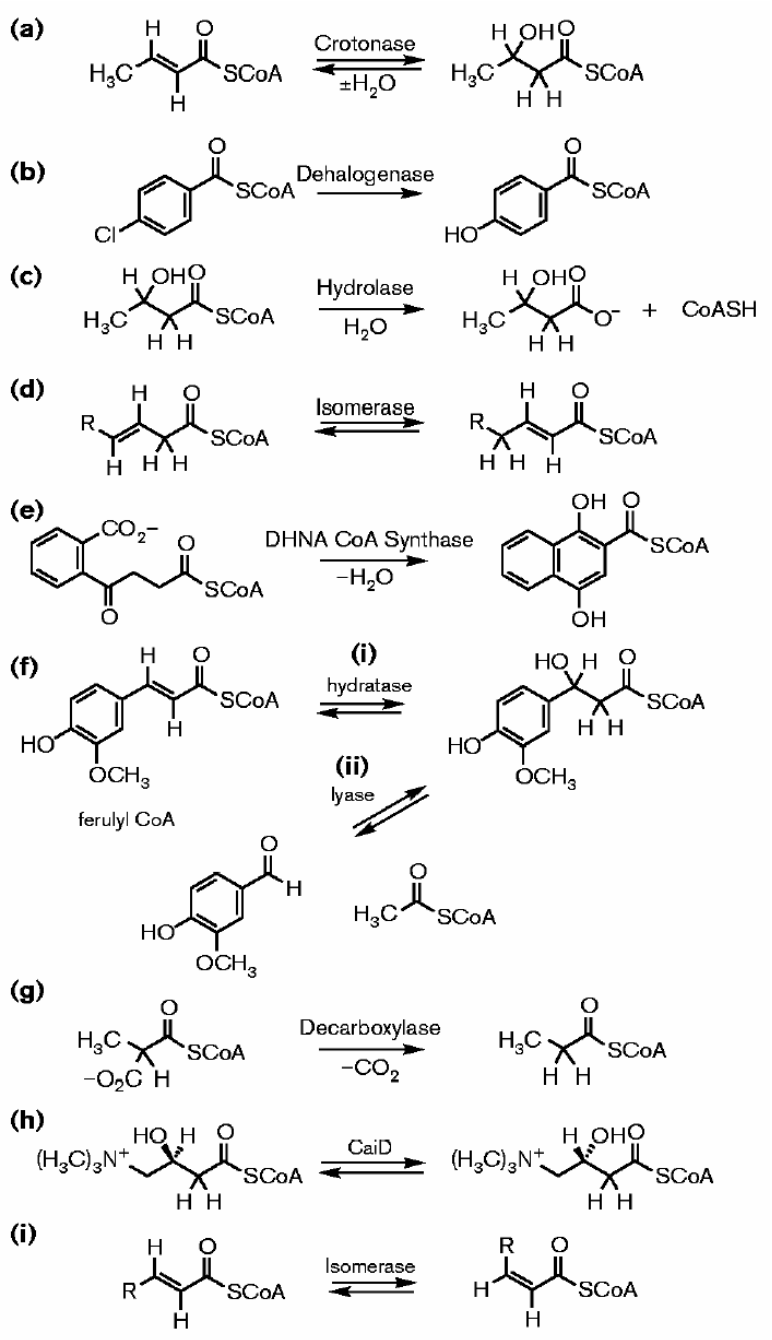


**Figure 11**. Reactions catalyzed by radical SAM superfamily members. Abbreviations: LAM, lysine 2,3-aminomutase; BioB, biotin synthase; LipA, lipoic acid synthase.

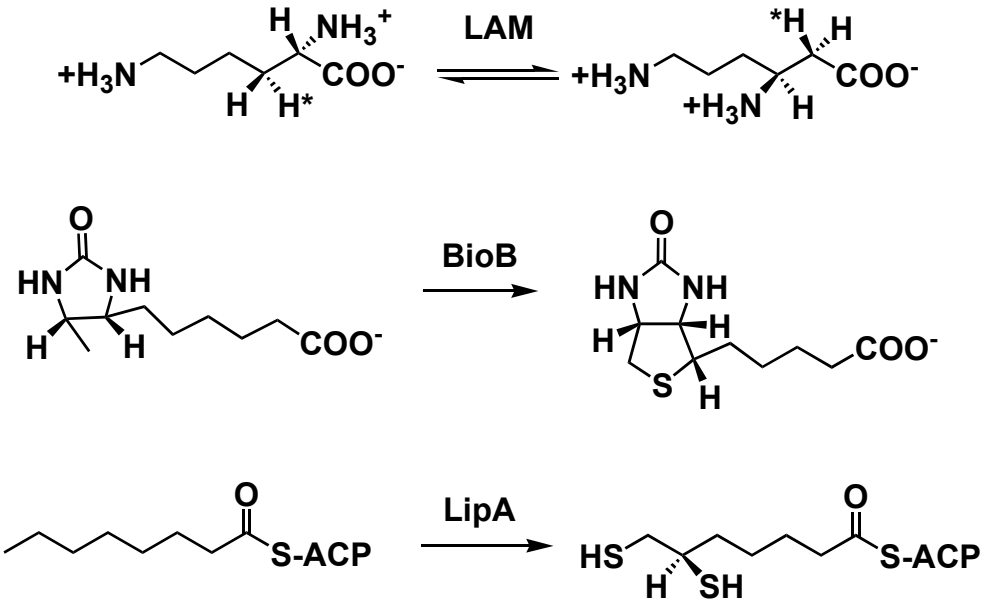


**Figure 12.** The catechol *meta*-fission pathway. Abbreviations: C2,3O, catechol 2,3-dioxygenase; 2-HMSD, 2-hydroxymuconate semialdehyde dehydrogenase; 4-OT, 4-oxalocrotonate tautomerase; 4-OD, 4-oxalocrotonate decarboxylase; VPH, vinylpyruvate hydratase; 2-HMSD, 2-hydroxymuconate semialdehyde hydrolase.

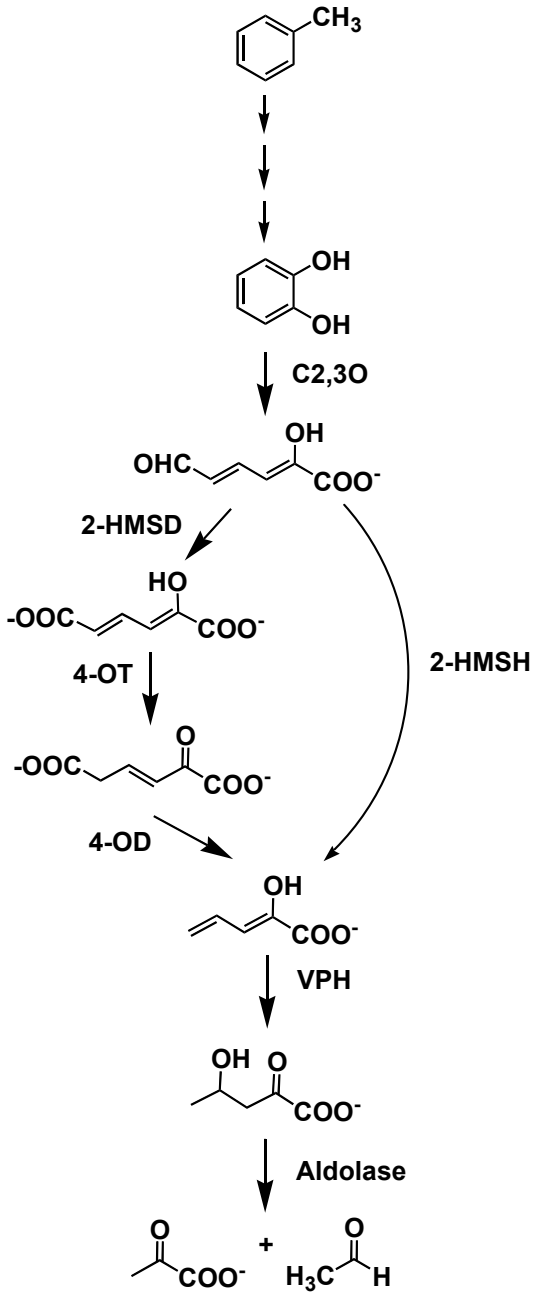


**Figure 13.** Sequence alignment of aldehyde dehydrogenases (ALDHs) generated using CLUSTAL-W. From top to bottom, betaine dehydrogenase, retinal dehydrogenase, 2-hydroxymuconate semialdehyde dehydrogenase, *Streptococcus mutans* ALDH, rat ALDH, *Vibrio harveyi* ALDH.

(*) indicates conserved residues. (-) indicates the Rossman fold. (+) and (¶) indicate residues for nucleotide binding. (‡) indicates an unconserved histidine implicated in catalysis. Symbols are located above the appropriate residue(s).

```
betaine   AQLVDSMPSASTGSV VVTDDLNYWGGRRIK SKDGATTEPVFEPAT GRVLCQMVPCGAEEV DQAVQSAQAAYLK-- -WSKMAGIERSRVML     87
retinal   -ASLQLLPSPTPN-- LEIKYTKIFINNEWQ NSESGRVFPVCNPAT GEQVCEVQEADKVDI DKAVQAARLAFSLGS VWRRMDASERGRLLD     88
hmsd      --------------- ---KEIKHFISGELV GSASGKLFDNVSPAN GQVIGRVHEAGRAEV DAAVRAARAALKG-- PWGKMTVAERAEILH     71
Strep     --------------- --TKQYKNYVNGEWK LSENE--IKIYEPAS GAELGSVPAMSTEEV DYVYASAKKAQPA-- -WRALSYIERAAYLH     69
rat       --------------- --------------- --------------- ------------SSI SDTVKRAREAFNS-- -GKTRSLQFRIQQLE     31
vibrio    --------------- NPQTDNVFYATNAFT GEALPLAFPVHTEVE VNQAATAAAKVARDF RRLNNSKRASLLR-- ------TIAS-ELEA     67
                                                                                                  β1
                                                                                                  ----    * ---
betaine   EAARIIRERRDNIAK LEVINNGK-TITEAE YDIDAAWQCIEYYAG LAPTLSGQHIQLPG- -----GAFAYTRREP LGVCAGILAWNYPFM    170
retinal   KLADLVERDRATLAT MESLNGGKPFLQAFY IDLQGVIKTLRYYAG WADKIHGMTIPVDG- -----DYFTFTRHEP IGVCGQIIPWNFPLL    172
hmsd      RVADGITARFGEFLE ARMPGHRQAEVAGQP HRHSARRANFKVFAD LLKNVANEAFEMATP DGA--GALNYGVRRP KGVIGVISPWNLPLL    159
Strep     KVADILMRDKEKIGA ILSKEVAK-GYKSAV SEVVRTAEIINYAAE EGLRMEGEVLEGGSF EAASKKKIAVVRREP VGLVLAISPFNYPVN    158
rat       ALQRMINENLKSISG ALASDLGKNEWTSYY EEVAHVLEELDTTIK ELPDWAEDEPVAKTR QTQ--QDDLYIHSEP LGVVLVIGAWNYPFN    119
vibrio    RSDDIIARAHLETAL PEVRLTGEIARTANQ LRLFADVVNSGSYHQ AILDTPNPTRAPLP- -----KPDIRRQQIA LGPVAVFGASNFPLA    151
             a1            β2           α2                  β3          α3         β4          α4
          ----------       ---+  ¶  ---------------         -----      -------     -----    ---------
betaine   IAAWK--CAPALACG NAVVFKPSPMTPVTG VILAEIFHEAG---- VPVGLVNVVQG--GA ETGSLLCHHPNVAKV SFTGSVPTGKKVMEM    252
retinal   MFTWK--IAPALCCG NTVVIKPAEQTPLSA LYMGALIKEAG---- FPPGVVNILPGY-GP TAGAAIASHIGIDKI AFTGSTEVGKLIQEA    255
hmsd      LMTWK--VGPALACG NCVVVKPSEETPLTA TLLGEVMQAAG---- VPAGVYNVVHGFGGD SAGAFLTEHPDVDAY TFTGETGTGETIMRA    243
Strep     LAGSK--IAPALIAG NVIAFKPPTQGSISG LLLAEAFAEAG---- LPAGVFNTITGR-GS EIGDYIVEHQAVNFI NFTGSTGIGERIGKM    241
rat       LTIQP--MVGAVAAG NAVILKPSEVSGHMA DLLATLIPQYM---- DQNLYLVVKGG---- VPETTELLKERFDHI MYTGSTAVGKIVMAA    199
vibrio    FSAAGGDTASALAAG CPVIVKGHTAHPGTS QIVAECIEQALKQEQ LPQAIFTLLQGN-QR ALGQALVSHPEIKAV GFTGSVGGGRALFNL    240
                 β5
          ---      ---*-                                  *
betaine   SAKT--VKHVTLELG GKSPLLIFKDCELEN AVRGALMANFLTQ-- GQVCTNGTRVFVQRE IMPQFLEEVVKRTKA IVVGDPLLT----ET    334
retinal   AGRSN-LKRVTLELG GKSPNIIFADADLDY AVEQAHQGVFFNQ-- GQCCTAGSRIFVEES IYEEFVKRSVERAKR RIVGSPFDP----TT    338
hmsd      AAKG--VRQVSLELG GKNAGIVFADCDMDK AIEGTLRSAFANC-- GQVCLGTERVYVERP IFDAFVARLKAGAEA LKIGEPNDP----EA    325
Strep     AGMR----PIMLELG GKDSAIVLEDADLEL TAKNIIAGAFGYS-- GQRCTAVKRVLVMES VADELVEKIREKVLA LTIGNPED-----DA    320
rat       AAKH--LTPVTLELG GKSPCYVDKDCDLDV ACRRIAWGKFMNS-- GQTCVAPDYILCDPS IQNQIVEKLKKSLKD FYGEDAKQ-----SR    280
vibrio    AHERPEPIPFYGELG AINPTFIFPSAMRAK ADLADQFVASMTMGC GQFCTKPGVVFALNT PETQAFIETAQSLIR QQSPSTLLTPGIRDS    330
                                                                                              *
betaine   RMGGLISKPQLDKVL GFVAQAKK-----EG ARVLCGGEPLTPSDP KLKNGYFMSPCVLDN CRDDMTCVKEEIFGP VMSVLPFDTEEEVLQ    419
retinal   EQGPQIDKKQYNKIL ELIQSGVA-----EG AKLECGGKGLG---- --RKGFFIEPTVFSN VTDDMRIAKEEIFGP VQEILRFKTMDEVIE    417
hmsd      NFGPLISHKPREKVP SYYQQAVD-----DG ATVVTGGGVPEMP-A HLAGGAWVQPTIWTG LADDSAVVTEEIFGP CCHIRPFDSEEEAIE    409
Strep     DITPLIDTKSADYVE GLINDAND-----KG ATALTEIKREG---- -----NLICPILFDK VTTDMRLAWEEPFGP VLPIIRVTSVEEAIE    396
rat       DYGRIINDRHFQRVK GLIDNQ--------- -KVAHGGTWDQ---- ---SSRYIAPTILVD VDPQSPVMQEEIFGP VMPIVCVRSLEEAIQ    353
vibrio    YQSQVVSRGSDDGID VTFSQAESPCVASAL FVTSSENWRKHPAWE EEIFGPQSLIVVCEN VADMLSLSEMEAGSL TATIHATEEDYPQVS    420
                                  ‡
betaine   RANNTTFGLASGVFT RDISRAHRVAANLEA GTCYINTYSIS--PV EVPFGGYKMSGFGRE NGQATVDYYSQLKTV IVEMGDVDSLF----    503
retinal   RANNSDFGLVAAVFT NDINKALMVSSAMQA GTVWINCYNAL--NA QSPFGGFKMSGNGRE MGEFGLREYSEVKTV TVKIPQKNS------    499
hmsd      LANSLPYGLASAIWT ENVRRAHRVAGQIEA GIVWVNSWFLR--DL RTAFGGSKQSGIGRE GGVHSLEFYTELKNI CVKL-----------    486
Strep     ISNKSEYGLQASIFT NDFPRAFGIAEQLEV GTVHINNKTQRG-TD NFPFLGAKKSGAGIQ GVKYSIEAMTTVKSV VFDIK----------    475
rat       FINQREKPLALYVFS NNEKVIKKMIAETSS GGVTANDVIVHITVP TLPFGGVGNSGMGAY HGKKSFETFSHRRSC LVKSL----------    433
vibrio    QLIPRLEEIAGRLVF NGWPTGVEVGYAMVH GGPYPASTHSASTSV GAEAIHRWLRPVAYQ ALPESLLPDSLKAEN PLEIARAVDGKAAHS    510
```

**Figure 14**. Proposed catalytic mechanism for aldehyde dehydrogenases.

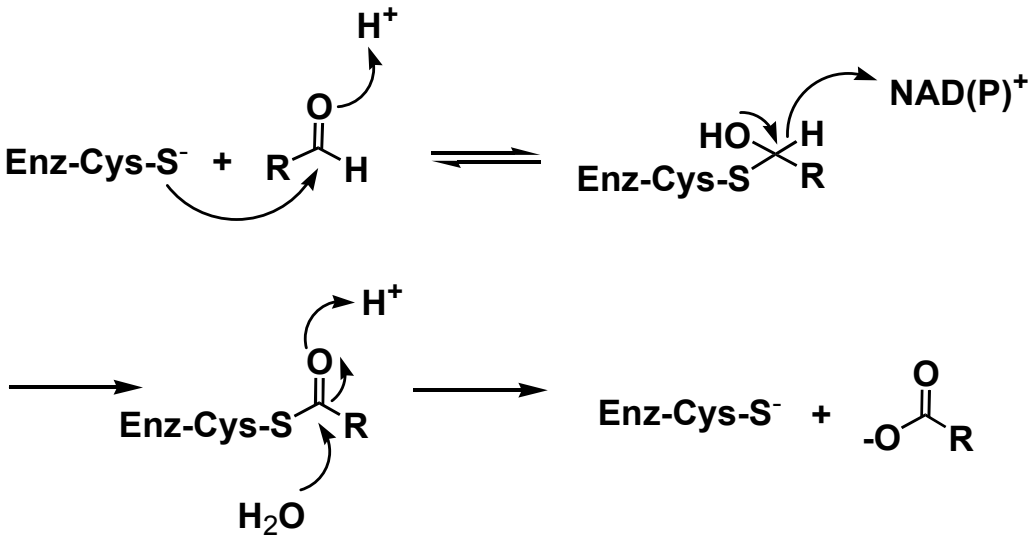


**Figure 15.** Active site residues of ALDHs (83). Abbreviations: Vh-ALDH, *Vibrio harveyi* ALDH; ALDH3, rat ALDH, ALDH2, mitochondrial ALDH, ALDH9, betaine dehydrogenase.

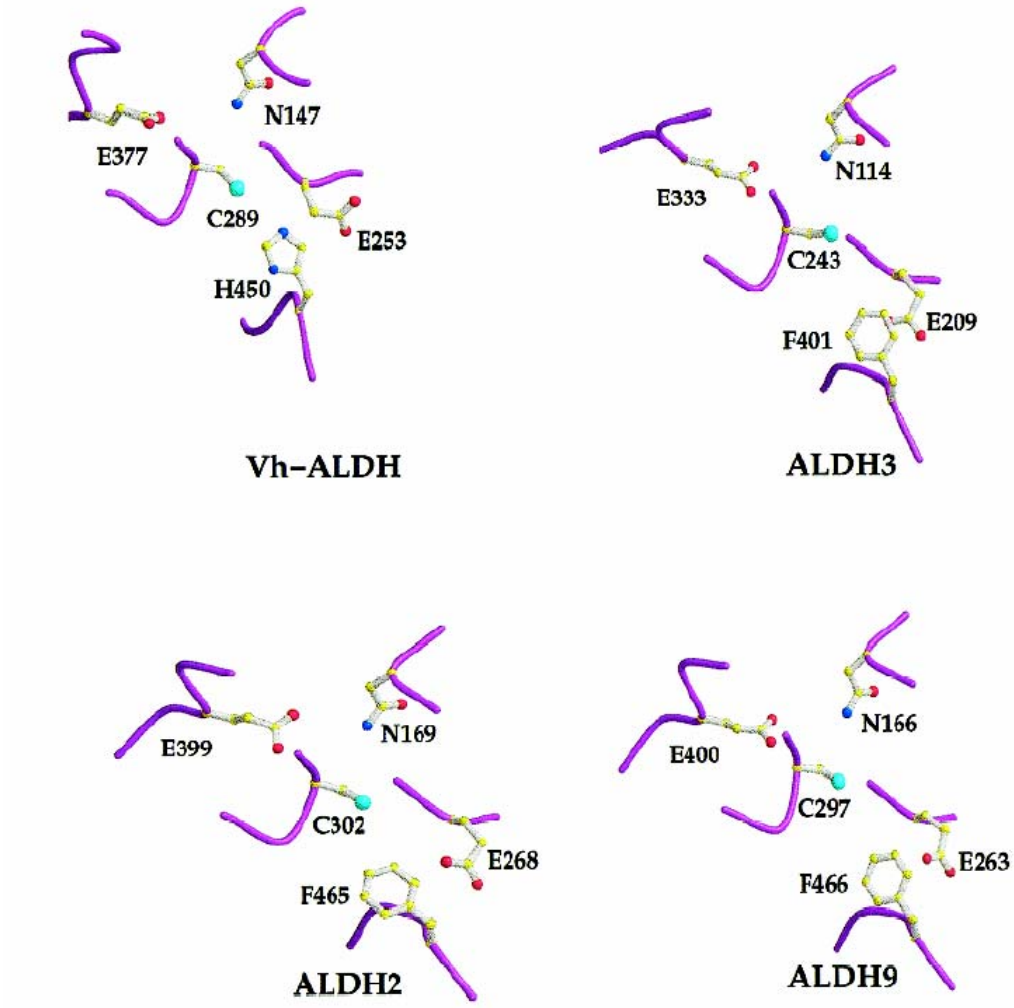


**Figure 16.** The 4-oxalocrotonate (4-OT)-catalyzed reaction.

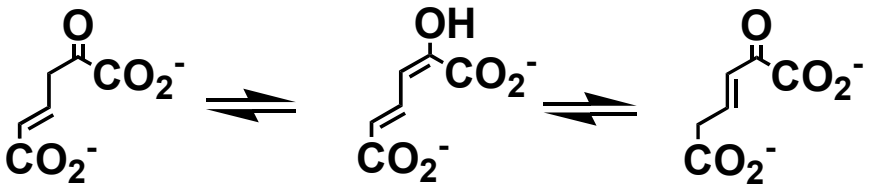


**Figure 17.** Proposed catalytic mechanism for 4-oxalocrotonate tautomerase (4-OT) (63).

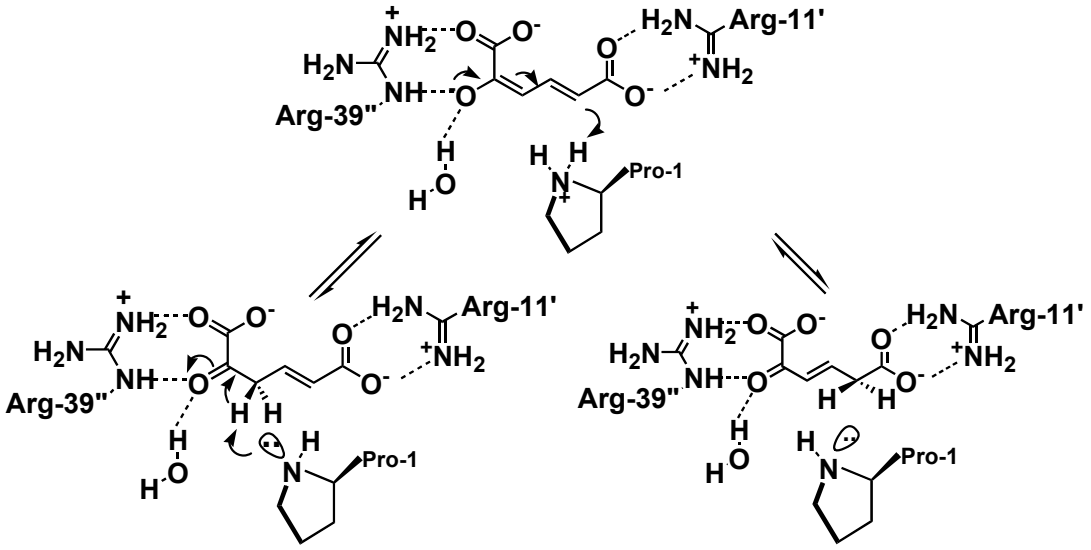


**Figure 18.** Reactions catalyzed by tautomerase superfamily members. Abbreviations, 4-OT, 4-oxalocrotonate tautomerase; MIF, macrophage migration inhibitory factor, CHMI, 5-carboxy-2-hydroxymuconate isomerase.

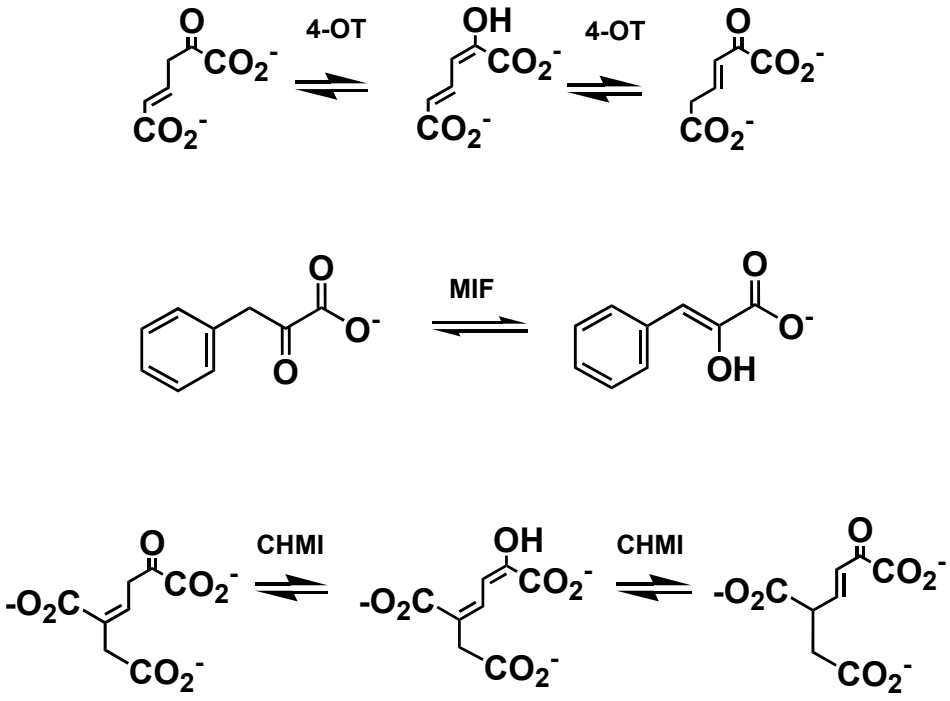


**Figure 19.** The homoprotocatechuate pathway for the degradation of aromatic amino acids. Abbreviations: HPCO, homoprotocatechuate (homogentisate) dioxygenase; CHMSD, 5-carboxy-2-hydroxymuconate semialdehyde dehydrogenase; CHMI, 5-carboxy-2-hydroxymuconate isomerase; COHED, 5-(carboxymethyl)-2-oxo-3-hexene-1,6-dioate decarboxylase; OHEDh, 2-oxo-hept-4-ene-1,7-dioate hydratase

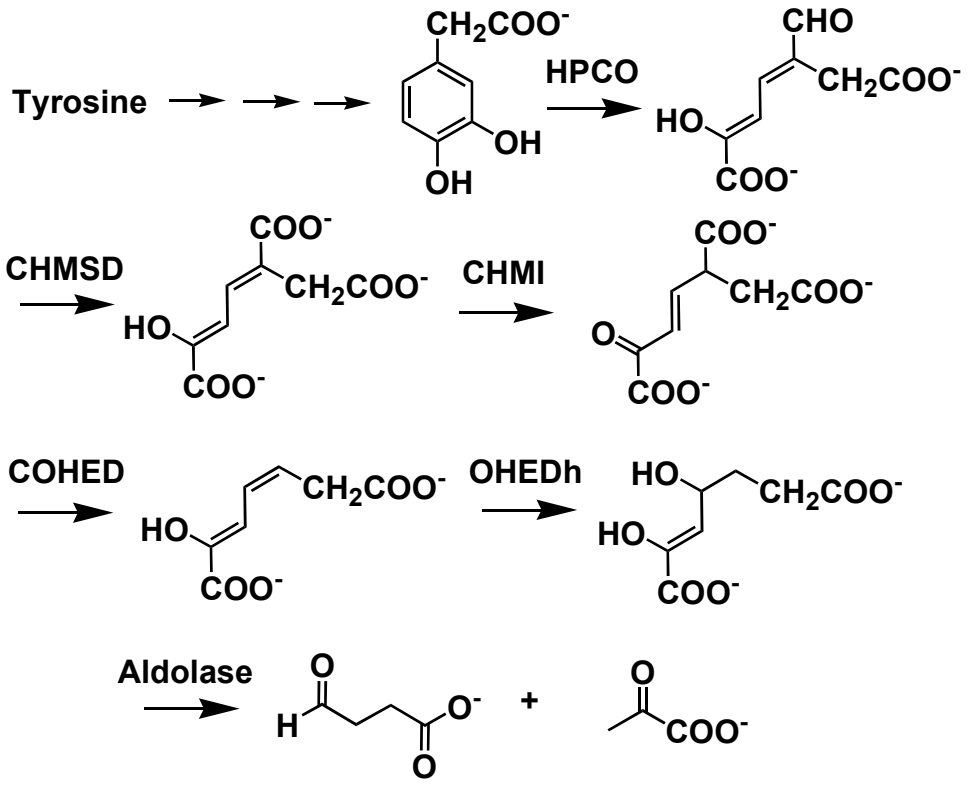


**Figure 20.** Sequence alignment of 4-oxalocrotonate tautomerase (4-OT), YwhB, and the α- and β-subunits of *trans*-3-chloroacrylic acid dehalogenase (CaaD) generated using CLUSTAL-W.

```
                1          11         21         31         41
4-OT            PIAQIHILEG RSDEQKETLI REVSEAISRS LDAPLTSVRV IITEMAKGHF
ywhB            PYVTVKMLEG RTDEQKRNLV EKVTEAVKET TGASEEKIVV FIEEMRKDHY
CaaDα           PMISCDMRYG RTDEQKRALS AGLLRVISEA TGEPRENIFF VIREGSGINF
CaaDβ           PFIECHIATG RSVARKQQLI RDVIDVTNKS IGSDPKIINV LLVEHAEANM

                51         61         71
4-OT            GIGGELASKV RR
ywhB            AVAGKRLSDM E
CaaDα           VEHGEHLPDY VPGNANDKAL IAKLK
CaaDβ           SISGRIHGEA ASTERTPAVS
```

**Figure 21**. Active site architectures of A) 4-OT and B) YwhB (63).

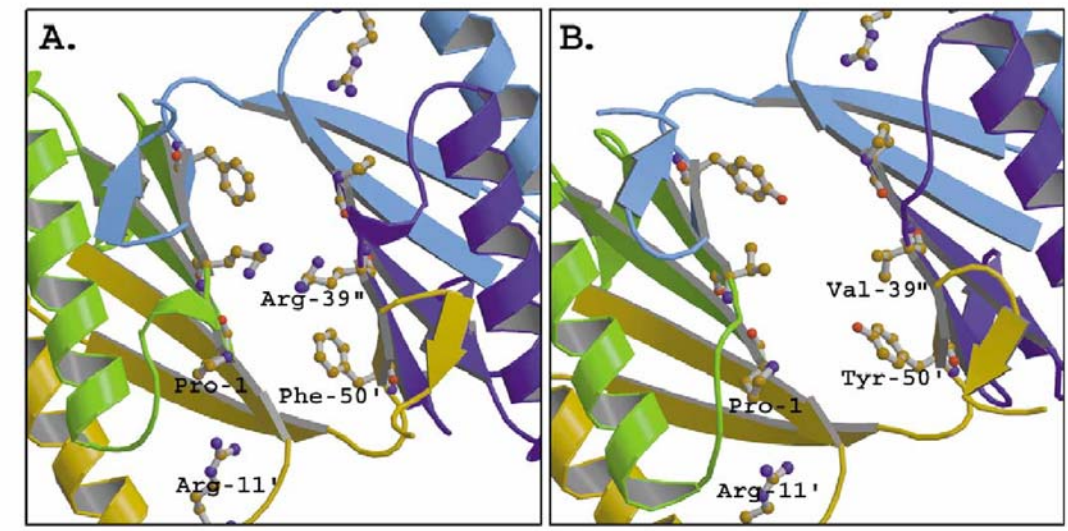


**Figure 22.** The reaction catalyzed by *trans*-3-chloroacrylic acid dehalogenase (CaaD).

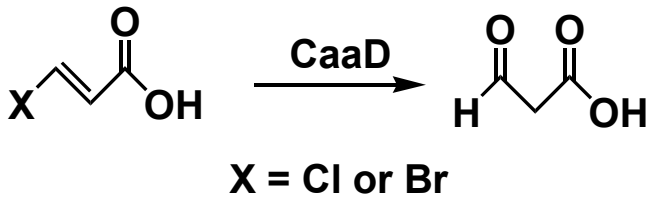
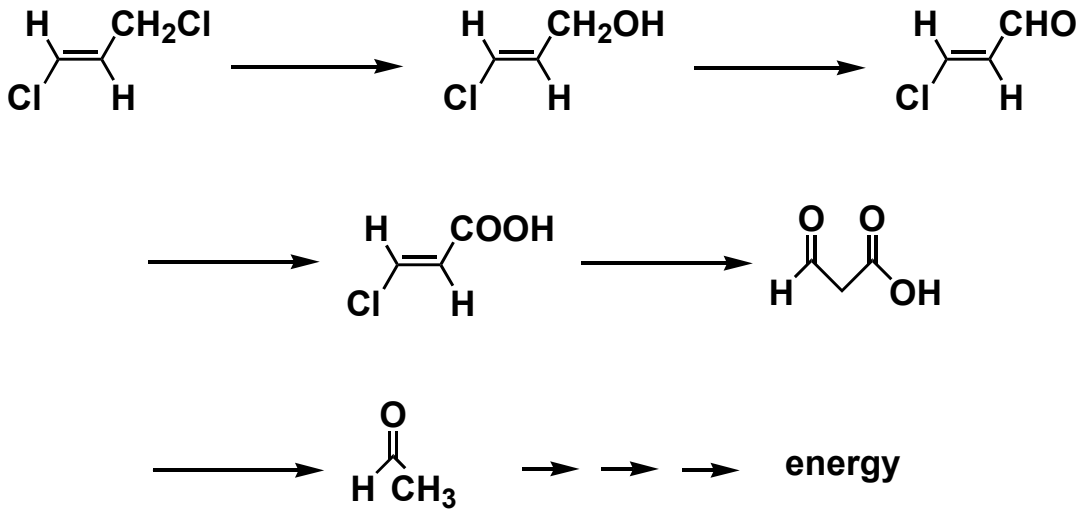


**Figure 23.** Proposed degradation pathway for the degradation of *trans*-1,3-dichloropropene (109).



# CHARACTERIZATION OF 2-HYDROXYMUCONATE SEMIALDEHYDE DEHYDROGENASE (2-HMSD)

## INTRODUCTION

2-Hydroxymuconate semialdehyde dehydrogenase (E.C. 1.2.1.32, 2-HMSD) from *Pseudomonas putida mt-2* catalyzes a reaction in the catechol *meta*-fission pathway (Scheme 1), oxidizing 2-hydroxymuconate semialdehyde (2-HMS, **1**) to 2-hydroxymuconate (2-HM, **2**, Scheme 2), requiring $NAD^+$ as a coenzyme (1). The enzymes of this pathway, encoded by the TOL plasmid pWW0, allow soil bacteria to use toluene, xylenes, and other substituted aromatics as their sole sources of carbon and energy. 2-HMSD is reportedly a homodimer of ~52 kDa subunits (2).

2-HMSD is a member of the semialdehyde dehydrogenase family within the aldehyde dehydrogenase (ALDH) superfamily of enzymes. ALDHs are ubiquitous among both prokaryotes and eukaryotes. They usually consist of ~500 amino acid monomers which assemble into homodimers or homotetramers in solution (3). ALDHs metabolize a vast array of aldehyde compounds which are often detrimental, causing cytotoxic reactions, cancer, and mutations (3). Thus, these enzymes serve an important detoxification role in biological systems. ALDHs also function in other cellular pathways such as glycolysis. A nucleotide coenzyme is an absolute requirement for activity, with most ALDHs exhibiting a strong preference (as assessed by $K_m$) for either $NAD^+$ or $NADP^+$.

On the basis of sequence alignment and crystallographic studies, a general catalytic mechanism can be formulated for ALDHs (Scheme 3). An active site cysteine attacks the aldehyde group, forming a thiohemiacetal intermediate. Subsequent hydride transfer to the nucleotide coenzyme occurs, resulting in oxidation of the thiohemiacetal and concomitant reduction of the $NAD(P)^+$ coenzyme. Deacylation of the enzyme leads to release of the carboxylate product. Cysteine is far more reactive as the thiolate ion; the corresponding p$K_a$ in the enzyme active site is

typically ~8.5 (4). Thus, another residue is thought to activate the cysteine by deprotonation. In the *Vibrio harveyi* ALDH, this residue is a histidine (5), while in others, such as rat ALDH, it is a glutamate (6-8). A conserved asparagine residue in the active site is thought to stablilize the thiohemiacetal intermediate through hydrogen bonding with the aldehyde oxygen (6, 9).

The sequence alignment of 2-HMSD with several ALDHs whose crystal structures have been solved is illustrated in Figure 24. The sequences shown represent the range of coenzyme and substrate specificity as well as a variety of host organisms. Four of these ALDHs, including 2-HMSD, are $NAD^+$-dependent. The first structure determined in the superfamily was that of class 3 rat liver ALDH, which is analogous to the human enzyme involved in alcohol metabolism. Betaine ALDH, found in humans and animals, converts betaine aldehyde to betaine, which can serve as a methyl donor in the biosynthesis of methionine and may also defend against osmotic stress (10, 11). Retinal dehydrogenase, found in various animal tissues, catalyzes the conversion of retinal to retinoic acid. Retinoic acid plays a variety of roles in differentiation and development, which are mediated through retinoid receptors. The two remaining ALDHs in the alignment are $NADP^+$-dependent. The ALDH from *Streptococcus mutans* catalyzes the irreversible oxidation of glyceraldehyde-3-phosphate into 3-phosphoglycerate, generating the NADPH necessary for growth and other biosynthetic pathways (12). Lastly, the ALDH from *Vibrio harveyi* catalyzes the oxidation of long-chain aliphatic aldehydes, substrates for this bacterium's chemiluminescence reaction (13).

Overall, the folds among the solved structures are quite similar despite differences in overall quaternary structure and the lack of a high degree of sequence identity. This alignment suggests the identities of potential catalytic residues in 2-HMSD. Putative active site residues in 2-HMSD include Asn-155, Cys-287, Glu-254, and Glu-390, indicated with an (*). The histidine in *Vibrio harveyi*, His-450, that deprotonates the cysteine is denoted with a (‡) (5). This residue is not conserved among the other ALDHs shown. In 2-HMSD, the corresponding residue is Ala-439.

A non-classical α/β Rossman fold, the canonical binding motif for nucleotide coenzymes, is shared by aldehyde dehydrogenases (Figure 24). The typical β-α-β Rossman fold (14) consists of six β-strands interconnected by five α-helices. In the ALDH Rossman fold, the α5 helix and the β6 strand are missing (8). The structure of the Rossman fold in 2-HMSD has not been determined.

Only one structure from the semialdehyde dehydrogenase family has been determined to date, that of aspartate β-semialdehyde dehydrogenase (ASADH) (15). However, ASADH catalyzes a reductive dephosphorylation that can be considered analogous to the reverse reaction of 2-HMSD (Scheme 4). Additionally, ASADH is an NADPH-dependent enzyme rather than an NADH-dependent one. Thus, it is likely that at least the nucleotide-binding domain of ASADH differs from that of 2-HMSD. ASADH and 2-HMSD do not share a high degree of sequence identity due to the much shorter length of the former enzyme (367 vs. 486 amino acids) and a disparity in the locations of the catalytic cysteines (Cys-135 in ASADH, Cys-287 in 2-HMSD). It would thus be of great interest to determine the structure of 2-HMSD, which would be the first $NAD^+$-dependent bacterial dehydrogenase structure available.

4-Oxalocrotonate tautomerase (4-OT) follows 2-HMSD in the *meta*-fission pathway (Scheme 1). It catalyzes a 1,3-keto-enol tautomerization as well as a 1,5-keto-enol tautomerization (Scheme 5). The overall reaction is the conversion of 2-oxo-4-hexenedioate (**3**) to 2-oxo-3-hexenedioate (**4**) via the dienol intermediate **2** (16). Since **2** exists in an equilibrium with two keto isomers in solution, it is likely that 2-HMS does as well (Scheme 6). The next enzyme in the pathway, 4-oxalocrotonate decarboxylase (4-OD) uses **4** as its substrate (Scheme 7) (Thanuja's 17). If 2-HMSD were to use the corresponding isomer of 2-HMS (i.e., **6**) as substrate, yielding **4** as product, 4-OT would be unnecessary in the pathway. Thus, it would be interesting to determine the isomeric equilibrium for **1** and to identify which isomer 2-HMSD processes.

The determination of the isomer of 2-HM generated by 2-HMSD would also identify the physiological substrate for 4-OT. This issue has been of interest in the Whitman laboratory for several years. Previous work has demonstrated that rapid equilibration occurs between **2** and **3**, followed by a slower conversion to **4** in aqueous solution (16). Kinetic and stereochemical studies indicate that both **2** and **3** are substrates for 4-OT, with **2** being the preferred substrate, as determined by $k_{cat}/K_m$ (16). Although **2** is stable in the crystalline form, a direct determination of the true substrate for 4-OT has been precluded by the instability of **3**, which cannot be synthesized or isolated.

In order to gain a better mechanistic understanding of 2-HMSD and to shed light upon its relationship to 4-OT, we first cloned the gene, then expressed and purified the protein. The gene for catechol 2,3-dioxygnease (C2,3O) was also cloned and the protein partially purified. C2,3O was used to generate 2-HMS from catechol via enzymatic synthesis (Scheme 8). The solution behavior of 2-HMS was examined by $^{1}H$ and $^{13}C$ NMR spectroscopy. Subsequently, we conducted a kinetic analysis of 2-HMSD with a variety of substrates (**1, 8-11**, Figure 25) and examined a potential inhibitor, the monoethyl ester of 2-HM (**12**, Figure 25). The role of the putative catalytic cysteine (Cys-287), identified through sequence analysis, was also investigated by irreversible inactivation experiments and site-directed mutagenesis.

In the course of this investigation, we determined that 2-HMS is a cyclic acetal when isolated by acidification and extraction (Scheme 9). NMR experiments indicate the resulting 6-membered ring (**14**) is stable in organic solvents but opens in aqueous solution to give **1**. The 2-HMSD-catalyzed physiological reaction (Scheme 2) yields **2** as product, suggesting that **2** is the cellular substrate for 4-OT. 2-HMSD also shows broad substrate specificity. Among the compounds investigated, 2-HMSD prefers 4-carboxybenzaldehyde (**11**), perhaps because its structure mimics the conformation of **1** that the enzyme requires (Figure 26). We also found that **12** is a competitive inhibitor of 2-HMSD, making it a useful ligand for future crystallographic studies. Furthermore, mutation of Cys-287 yielded inactive (C287A)

or insoluble (C287S) protein. These results, coupled with time-dependent irreversible inactivation using iodoacetamide (**15**), indicate that Cys-287 is required for activity, and suggest that it may be a key catalytic residue.

## MATERIALS AND METHODS

*Materials.* Tryptone, yeast extract, and agar were obtained from Difco (Detroit, MI) or Fisher Scientific (Pittsburgh, PA). Restriction enzymes, PCR reagents, Taq DNA polymerase, Platinum *Taq* High Fidelity DNA polymerase and T4 DNA ligase were obtained from F. Hoffmann-La Roche, Ltd. (Basel, Switzerland) or Promega Corp. (Madison, WI). *Pfu* DNA polymerase was obtained from Stratagene (La Jolla, CA). Thin-walled PCR tubes were obtained from Ambion (Austin, TX). Oligonucleotides for DNA amplification and sequencing were synthesized by either Genosys (The Woodlands, TX) or Oligos Etc. (Wilsonville, OR). The Wizard Plus Minipreps and PCR Preps DNA Purification Systems were obtained from Promega. The remaining molecular biology reagents including agarose, DNA ladders, and protein molecular weight standards, were obtained from Invitrogen Corporation (Carlsbad, CA). The 0.2 μM PES syringe filters were obtained from Nalge Nunc International (Rochester, NY). The Amicon stirred cells and ultrafiltration membranes were obtained from Millipore Corp. (Billerica, MA). The pET-24a(+) plasmid was obtained from Novagen (Madison, WI). All other reagents were obtained from Sigma-Aldrich, Inc. (St. Louis, MO).

*Strains*. *Escherichia coli* strain DH5α (Invitrogen) was used for cloning and isolation of plasmids. *E. coli* strains BL21(DE3)pLysS (Novagen) and BL21-Gold(DE3)pLysS (Stratagene) were used for recombinant protein expression.

*Methods*. Techniques for restriction enzyme digestions, ligation, transformation, and other standard molecular biology manipulations were based on methods described elsewhere (18), or as suggested by the manufacturer. Plasmid DNA was introduced into cells by electroporation using a Cell-Porator Electroporation System (Whatman Biometra, Göttingen, Germany). The PCR was performed in a DNA thermal cycler (Model 480) obtained from PerkinElmer Inc. (Wellesley, MA). Colony screening by the PCR is based on methods described in the pET System Manual (Novagen). DNA sequencing was done at the DNA Core

Facility in the Institute for Cellular and Molecular Biology at the University of Texas (Austin). Kinetic data were obtained on a Hewlett Packard 8452A or an Agilent 8453 Diode Array spectrophotometer. Solutions in cuvettes were mixed using a stir/add cuvette mixer (Bel-Art Products, Pequannock, NJ). The kinetic data were fitted by nonlinear regression data analysis using the Grafit program (Erithacus Software Ltd., Horley, U. K.) obtained from Sigma Chemical Co. HPLC was performed on a Waters (Milford, MA) 501/510 system or a Beckman System Gold HPLC (Fullerton, CA) using a TSKgel Phenyl-5PW hydrophobic column or a TSKgel DEAE-5PW anion exchange column (Tosoh Biosciences, Montgomeryville, PA). Protein was analyzed by Tris glycine sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions on 10-17.5% gels using either the Mini-Protean II or III vertical gel electrophoresis apparatus obtained from Bio-Rad (Hercules, CA) (19). Protein concentrations were determined using the method of Waddell (20). The NMR spectra were recorded in 100% $H_2O$ on a Varian Unity INOVA-500 spectrometer using selective pre-saturation of the water signal with a 2-s pre-saturation interval. The lock signal is dimethyl-$d_6$ sulfoxide. Chemical shifts are standardized to the dimethyl-$d_6$ sulfoxide signal at 2.49 ppm.

*Construction of the C2,3O expression vector.* Catechol 2,3-dioxygenase has previously been cloned from *Pseudomonas putida* mt-2, and its structure has been determined (21, 22). In order to synthesize **1**, the *xylE* gene was cloned from *Pseudomonas putida mt-2* using overlap extension PCR (23) to remove an internal *Nde*I restriction site. The TOL plasmid was prepared as described elsewhere (17). The external primers were oligonucleotides 5'-GGT ATT CCA TAT GAA CAA AGG TGT AAT G-3' (primer A) and 5'-CGG AAT TCT CAG GTC AGC ACG GT-3' (primer D). Primer A contains 7 bases (to facilitate digestion), an *Nde*I restriction site (underlined), and 15 bases corresponding to the nucleotide sequence of the first five amino acids of C2,3O. Primer D contains 9 bases (to facilitate digestion), an *Eco*RI restriction site (underlined), and 8 bases corresponding to the reverse complement of the C2,3O nucleotide sequence. The internal primers were 5'-ACA

GCC ATA AGC CAT CAG ATC-3' (primer B) and 5'-GAT CTG ATG GCT TAT GGC TGT-3' (primer C). In each internal primer, the mutation (a single base) is underlined and remaining bases correspond to the coding sequence (primer C) or the complementary sequence (primer B) of the gene.

For the construction of the AB and CD fragments, each PCR reaction contained 5 μL of 20 μM solutions of primers A and B or C and D, 2 μL of a 10 mM stock of dNTPs, 10 μL of 10x *Taq* reaction buffer, 6 μL of 25 mM $MgCl_2$, 8 μL of template, 1 μL of *Taq*, and a sufficient quantity of sterile water to make a final volume of 100 μL. The PCR protocol consisted of a 2-min incubation period at 94 °C, followed by 20 cycles of the PCR where each cycle contained a denaturation step at 94 °C for 2 min, an annealing step at 50 °C for 2 min, and an elongation step at 72 °C for 3 min, and a final 5-min incubation period at 72 °C. The amplified products were analyzed by electrophoresis on 0.8% agarose gels and purified using the Wizard PCR Preps system. For the construction of the AD fragment, each PCR reaction was assembled similarly except that 5 μL of each of the AB and CD fragments were used as template. The PCR protocol consisted of a 2-min incubation period at 94 °C, followed by 25 cycles of the PCR where each cycle contained a denaturation step at 94 °C for 1 min, an annealing step at 55 °C for 75 s, and an elongation step at 72 °C for 75 s, and a final 10-min incubation period at 72 °C. The AD fragment was electrophoresed. Two bands (~850 and ~950 bp) were observed, which were both excised and purified using the Wizard PCR Preps system.

Subsequently, the purified PCR products and the pET-24a(+) vector were treated with the *Nde*I and *Eco*RI restriction enzymes for 3 hrs at 37° C. The digestion products were electrophoresed on a 0.8% agarose gel. Further purification was achieved by use of the Wizard PCR Preps system. The resulting fragments were ligated using T4 ligase (0.1 μL of 10 units/μL) at room temperature overnight. The DNA was precipitated from the mixture and re-suspended in sterile water (5 μL), which was used to transform *E. coli* strain DH5α cells by electroporation according to the manufacturer's instructions. The transformed cells were treated as previously

described (21), plated on LB/Kn (50 μg/mL) agar plates, and grown overnight at 37 °C. Growth resulted on plates containing the larger (i.e., ~950 bp) C2,3O fragment. Colony screening was performed on 20 colonies to detect the presence of insert. Four colonies were selected for plasmid isolation and sequencing. Only one of the four plasmids contained the correct sequence. This plasmid was designated pET-24a(+)-C2,3OA.

Aliquots (0.5-1 μL) of pET-24a(+)-C2,3OA were used to transform either *E. coli* strain BL21(DE3)pLysS or *E. coli* strain BL21-Gold(DE3)pLysS cells by electroporation following the manufacturer's directions. An aliquot (50 μL) was plated on LB/Kn (50 μg/mL) agar plates, and grown overnight at 37 °C. In the presence of IPTG (0.5 mM) and catechol (μM), cells expressing C2,3O gave rise to a yellow color, indicating the production of 2-HMS (24).

*Overexpression and Partial Purification of C2,3O*. A 50 mL culture of LB containing kanamycin (50 μg/mL) was inoculated with a single colony of the expression strain and grown overnight at 37° C. Subsequently, a sufficient quantity of the overnight culture was used to inoculate 500 mL of LB/Kn medium in a 2 L Erlenmeyer flask to give an $OD_{600nm}$ reading of ~0.05. Cultures (6-2L flasks) were grown at 37° C to an $OD_{600nm}$ of ~0.6-0.8 with vigorous shaking and induced with 0.5 mM IPTG. Cells were harvested after three hours of growth and stored at -80° C until use. Typically, 2 L of culture yields 5 g cells.

C2,3O was partially purified using a published protocol modified as follows (25). The cells (5 g) were thawed, resuspended in 3 volumes (14 mL) of 10 mM ethylenediamine buffer containing 10% isopropanol (pH 7.4), placed on ice, and initially disrupted by sonication at 50% output for 30 s. Phenylmethylsulfonyl fluoride (0.5 mM) and 6-aminocaproic acid (1 mM) were added to the cell suspension. Subsequently, the cell suspension was subjected to further sonication for four 5-min intervals using 5-s pulses with a cycle time of 50%. After each 5-min interval, the cells were allowed to sit for 5 minutes. Protamine sulfate (2% of the cell weight, 96 mg) was added to the lysate, and the resulting mixture was stirred for 10

min at 4 °C and centrifuged at 19,000 rpm for 30 min at 4 °C. The supernatant was filtered through a 0.2 μM PES syringe filter.

The filtered supernatant was loaded onto a DEAE-5PW anion-exchange column (21.5 mm x 15 cm) previously equilibrated in 10 mM ethylenediamine buffer, pH 7.4 (Buffer A). Protein was eluted using a 10 min wash of 100% Buffer A, followed by 2-25 min linear sodium sulfate gradients (0-0.15 M, 0.15-0.25 M), and a final 10-min wash with 0.5 M sodium sulfate. The flow rate was 5 mL/min. Fractions (7.5-10 mL) were collected and assayed for activity. Fractions were assayed for activity using a mixture containing 2 μL catechol (165 mM in EtOH), 5 μL of the fraction, and 1 mL of 100 mM sodium phosphate buffer, pH 7.3. The increase in absorbance at 375 nM, due to the formation of **1**, was monitored. Typically, the C2,3O activity elutes at ~0.17 M sodium sulfate. The most active fractions were pooled (~70 mL), concentrated, and exchanged into 20 mM sodium phosphate (pH 7.3) using an Amicon stirred cell containing a YM-10 membrane. This purification resulted in ~150 mg of ~75% pure protein as assessed by SDS-PAGE.

*Construction of the 2-HMSD expression vector*. The *xylG* gene from *Pseudomonas putida* mt-2 was cloned in two stages using a modification of overlap extension PCR (23). DNA from the TOL plasmid was isolated using the Wizard Plus Minipreps. The external primers used were 5'-AGA GAT TGT GCA TAT GAA AGA AAT CAA GCA T-3' (primer A) and 5'-GTT CAT GAG CGG CCG CCT CAA AGT TTC ACA CAG AT-3' (primer D). Primer A contains 10 bases (to facilitate digestion), an *Nde*I restriction site (underlined), and 15 bases corresponding to the coding nucleotide sequence of the first five amino acids. Primer D contains 7 bases (to facilitate digestion), a *Not*I restriction site, 1 additional base, and 18 bases corresponding to the reverse complement of a stop codon and the nucleotide sequence for the last five amino acids of 2-HMSD. The internal primers were 5'-GCC TGC ATC ACC TCG CCG A-3' (primer B) and 5'-GGT CAA ACC ATC CGA GGA A-3' (primer C). These primers correspond to the nucleotide coding sequence of 2-HMSD

including the last base of Val-177 through Glu-183, and the reverse complement of the sequence including the last base of Leu-189 through Gln-195, respectively.

For the construction of the AB and CD fragments, each PCR reaction contained 2 μL of 20 μM solutions of primers A and B or C and D, 2 μL of a 10 mM stock of dNTPs, 10 μL of *Pfu* reaction buffer, 15 μL of the TOL plasmid, 2 μL of *Pfu*, and a sufficient quantity of sterile water to make a final volume of 100 μL. The PCR protocol consisted of a 45-s incubation period at 94 °C, followed by 25 cycles of the PCR in which each cycle contained a denaturation step at 94 °C for 45 s, an annealing step at 55 °C for 45 s, and an elongation step at 72° C for 3.5 min, and a final 10 min incubation period at 72 °C. The amplified products were analyzed by electrophoresis on 0.8% agarose gels and purified using the Wizard PCR Preps system. The AD fragment was constructed similarly except that the AB and CD fragments (5 μL of each) were used as template, and the 72 °C elongation step for each cycle was increased to 4 min. A band of the anticipated size (~1500 bp) was excised after analysis by gel electrophoresis and purified using the Wizard PCR Preps.

The purified AD fragment and the pET-24a(+) vector were treated with *Nde*I and *Not*I overnight at 37 °C. The digestion products were electrophoresed and purified using the Wizard PCR Preps. Ligation and precipitation of the DNA were carried out as described previously. The purified DNA was used to transform *E. coli* strain DH5α by electroporation following the manufacturer's instructions. The transformed cells were treated and incubated as described above. Colony screening was performed on four colonies to detect the presence of insert. Subsequent sequencing of one plasmid, designated pET-24a-2-HMSD 8, indicated differences from the published sequence (Genbank Accession No. CAC86805). This plasmid was used for all further studies. Preparation of the expression strain (BL21(DE3)pLysS or BL21-Gold(DE3)pLysS) was performed as described above.

*Overexpression and Purification of 2-HMSD.* The expression strain was grown under the conditions described above. Typically 3L of culture yields 6-7 g of cells. 2-HMSD was purified by modifying a series of previously published

procedures (1, 2, 26). The buffers used in this purification are Buffer A: 50 mM $KH_2PO_4$ buffer containing 10% isopropanol, pH 8.2; Buffer B: 50 mM $KH_2PO_4$ buffer made 1 M in $(NH_4)_2SO_4$, pH 7.4; and Buffer C: 50 mM $KH_2PO_4$ buffer, pH 7.4. The cells (6 g) were thawed, resuspended in 3 volumes (18 mL) of buffer A, placed on ice, and initially disrupted by sonication at 50% output for 30 s. Phenylmethylsulfonyl fluoride (0.5 mM) and 6-aminocaproic acid (1 mM) were added to the cell suspension. Subsequently, the cell suspension was subjected to further sonication for three 5-min intervals using 5-s pulses with a cycle time of 50%. After each 5-min interval, the cells were allowed to sit for 5 minutes. Protamine sulfate (2% of the cell weight, 0.12 g) was added to the lysate. The mixture was stirred for 10 min at 4 °C and centrifuged at 17,000 rpm for 30 min at 4 °C.

The resulting supernatant was filtered through a 0.2 μm syringe filter and loaded onto a DEAE-5PW anion-exchange column previously equilibrated in Buffer A. Protein was eluted using a 10-min wash of 100% Buffer A, followed by a 50 min linear gradient (0-1 M) in sodium chloride, and a final 10-min wash with 1 M NaCl. The flow rate was 5 mL/min. Fractions (7.5-10 mL) were collected and assayed for activity. Typically, 2-HMSD elutes at ~0.25 M NaCl. The most active fractions were pooled (~60 mL) and concentrated using an Amicon stirred cell equipped with a YM-30 membrane. The resulting concentrate (~10 mL) was made 1 M in $(NH_4)_2SO_4$ and allowed to stir for 1 hr at 4 °C. After centrifugation at 15,000 rpm for 30 min, the supernatant was removed and loaded onto a Phenyl-5PW hydrophobic column (21.5 mm x 15 cm) equilibrated in Buffer B. Protein was eluted using a 10 min wash of 100% Buffer B, followed by a 30 min linear gradient (1-0 M) using Buffer C, and a final 20-min wash with 100% Buffer C. Fractions were assayed and concentrated as before. 2-HMSD elutes during the final wash. The purity was judged to be ~99% as assessed by SDS-PAGE. The overall yield is approximately 15-30 mg of homogenous protein per liter of culture. For this preparation, a 3-L culture gave 90 mg of protein.

*Construction of the C287A and C287S 2-HMSD Mutants.* Two mutants were constructed by overlap extension PCR as described elsewhere (23). Sequence alignment shows that a single cysteine is conserved at position 287. Thus, Cys-287 was mutated to an alanine and a serine. The external primers for the construction of the alanine mutant were primers A and D as described in the construction of the 2-HMSD expression vector. The internal primers were 5′ - TCG GTG CCC AGT GCG ACC TGT CC - 3′ (primer B) and 5′ - GGA CAG GTC GCA CTG GGC ACC GA - 3′ (primer C). In each internal primer, the mutation is underlined and the remaining bases correspond to the coding sequence (primer C) or the complementary sequence (primer B) of 2-HMSD.

For the construction of the AB and CD fragments for the C287A mutant, each PCR reaction contained 1 μL of 20 μM solutions of primers A and B or C and D, 1 μL of a 10 mM stock of dNTPs, 5 μL of 10 × Platinum *Taq* reaction buffer, 2 μL of 50 mM $MgSO_4$, 1 μL of pET-24a-2-HMSD8, 0.4 μL Platinum *Taq* High Fidelity, and a sufficient quantity of sterile water to make a final volume of 50 μL. The PCR protocol consisted of a 45-s incubation period at 94 °C, followed by 35 cycles of the PCR where each cycle contained a denaturation step at 94 °C for 30 s, an annealing step at 50 °C for 45 s, and an elongation step at 68 °C for 2 min, and a final 10-min incubation period at 68 °C. The amplified products were analyzed by electrophoresis on 0.8% agarose gels and purified using the Wizard PCR Preps DNA purification system.

The AD fragment was constructed similarly except that the AB and CD fragments (2 μL of each) were used as template. The AD fragment was electrophoresed, excised, and purified using the Wizard PCR Preps system. Restriction digests and subsequent DNA purification were performed as described previously. The resulting fragments were ligated using T4 ligase (0.1 μL of 10 units/μL) at room temperature for 1 hr. An aliquot (1 μL) of the ligation was diluted five-fold with sterile water and used to transform *E. coli* strain DH5α cells by electroporation. The transformed cells were treated as previously described (21),

plated on LB/Kn (50 μg/mL) agar plates, and grown overnight at 37 °C. Plasmid DNA was isolated from the single colony that resulted. Sequencing of the surrounding bases (~400 bp) indicated that the mutation was present, and that no other mutations had been introduced in that section of the gene. This plasmid was designated pET-24a-2-HMSD-C287A.

The external primers for the C287S mutant were oligonucleotides 5'-GCG GAT AAC AAT TCC CCT CT-3' (primer A) and 5'-CTC AGC TTC CTT TCG GGC TT-3' (primer D). Primer A corresponds to a region 36 bases upstream of the *Nde*I restriction site in the pET-24a(+) plasmid, while primer D corresponds to the complementary sequence of a region 20 bases downstream of the His-tag region in the pET-24a(+) plasmid. The internal primers were primer B from the C287A mutant and oligonucleotide 5′ - GGA CAG GTC AGT CTG GGC ACC GA - 3′ (primer C). In primer C, the mutation is underlined and the remaining bases correspond to the coding sequence of the gene.

For the construction of the AB and CD fragments for the C287S mutant, each PCR reaction contained 1 μL of 20 μM solutions of primers A and B or C and D, 1 μL of a 10 mM stock of dNTPs, 5 μL of 10 × *Pfu* buffer, 1 μL of pET-24a-2-HMSD8, 1 μL of *Pfu* DNA Polymerase, and a sufficient quantity of sterile water to make a final volume of 50 μL. The PCR protocol used was the same as that used for the synthesis of the AB and CD fragments for wild-type 2-HMSD. The amplified products were analyzed by electrophoresis on 0.8% agarose gels. Two bands (~1 kb) were observed from the AB reaction; a single band (~500 bp) was observed from the CD reaction. Each band was purified using the Wizard PCR Preps DNA purification system. The AD fragment was constructed similarly except that the AB and CD fragments (2.5 μL of each) were used as template, and the same PCR protocol was followed. A single band (~1500 bp) was observed after electrophoresis, which was excised and purified using the Wizard PCR Preps system.

Restriction digests and subsequent purification of the pET-24a(+) vector and the AD fragment were performed as described previously. The resulting fragments

were ligated using T4 ligase (0.1 μL of 10 units/μL) at room temperature for 2 hr. An aliquot (2 μL) of the ligation mixture was diluted five-fold with sterile water and used to transform *E. coli* strain DH5α cells by electroporation. The transformed cells were treated as previously described (21), plated on LB/Kn (50 μg/mL) agar plates, and grown overnight at 37 °C. Two colonies resulted, from which plasmid DNA was isolated. Sequencing of the surrounding bases (~400 bp) indicated that the mutation was present in both plasmids, and that no other mutations had been introduced in that portion of the gene. These plasmids were designated pET-24a-2-HMSD C287S 1 and pET-24a-2-HMSD C287S 2.

*Overexpression of the C287A and C287S 2-HMSD Mutants.* For expression, each mutant plasmid was transformed into BL21(DE3)pLysS by electroporation. Cells were selected for as described previously and grown using the protocol described for wild-type 2-HMSD. Subsequently, the C287A mutant was purified using the procedure used for wild-type 2-HMSD. One liter of culture yields 2 g of cells and 15 mg of protein. Approximately half of the induced protein is found in the initial pellet after cell lysis. One liter of C287S HMSD culture yields 2 g of cells. However, all of the induced protein was found in the cell pellet. Therefore, no further attempt at purification was pursued.

*Synthesis of 2-HMS* (**1**). Compound **1** was generated enzymatically using C2,3O as follows. Catechol (200 mg, 1.8 mmol) was dissolved in 20 mM sodium phosphate buffer (30 mL, pH 7.5). To this mixture was added a 100 μL aliquot of C2,3O (11.5 mg/mL). Pure oxygen gas was bubbled directly into the reaction as it stirred. Subsequently, aliquots (50 μL) of C2,3O were added every 15 min (4x) and the pH adjusted to 7.3-7.6 with aliquots of 1 M NaOH. When the pH no longer required adjustment, indicating that the acid was no longer being generated, the reaction was complete. As the reaction progressed, a yellow color appeared which turned to deep red, indicative of the formation of **1**. The solution was acidified to pH ~2 with 8.5% phosphoric acid and extracted with ethyl acetate (3x30 mL). The organic layers were pooled, dried over anhydrous $Na_2SO_4$, filtered, and evaporated to

dryness at room temperature. The resulting solid was dissolved in ethyl acetate and recrystallized. The compound is stable at room temperature as a solid. A $^1$H NMR spectrum taken in deuterated methanol indicates that compound **14** is isolated. However, the $^1$H NMR spectrum indicates that compound **1** exists in aqueous solution. The yield of **14** was 48 mg (0.3 mmol, 17%).

*Enzymatic Assays and Kinetic Studies of 2-HMSD.* The kinetic assays were performed in 50 mM $KH_2PO_4$ buffer, pH 8.5 using **1** or **8-11** and $NAD^+$. Stock solutions of $NAD^+$ (0.1 M) were made in buffer (50 mM $KH_2PO_4$, pH 8.5) and stored at 4 °C. The solutions were stable for a week under these conditions. Solutions of $NAD^+$ (5, 25, and 50 mM) were made by diluting the stock solution appropriately into 50 mM $KH_2PO_4$ buffer, pH 8.5. Stock solutions of **1** (20 mM), **9** (200 mM), **10** (200 mM), and **11** (20 mM) were prepared in ethanol. Solutions of **1** and **11** (2 mM) and **10** (100 mM) were made by appropriate dilution of stock solutions with ethanol. Stock solutions (154 mM) of **8** were prepared in 50% DMSO/50% water (v/v).

To measure the kinetic parameters for compounds **1** and **8-11**, a sufficient quantity of 2-HMSD was diluted into buffer (40 mL). $NAD^+$ was added to the incubation mix at a saturating condition (1 mM, 10 μL of a 0.1 M solution) as described elsewhere (1, 2, 26). The enzyme/coenzyme mixture was equilibrated at 23 °C for at least 15 min prior to assaying. The quantities of enzyme used with the indicated compounds are as follows, diluted from a 3.0 mg/mL solution of 2-HMSD: **1**, 80 μL/40 mL; **8** and **9**, 120 μL/40 mL; **10**, 800 μL/40 mL; **11**, 80 μL/40 mL. The final concentrations of enzyme ranged from $10^{-8}$-$10^{-6}$ M. Subsequently, a 1 mL-portion of the mixture was transferred to a cuvette and assayed by the addition of a small quantity of substrate. For **1**, the decrease in absorbance at 375 nm ($\varepsilon_{2\text{-HMS}}$ = 33,000 $M^{-1}$ $cm^{-1}$), corresponding to the decrease in the concentration of substrate, was monitored. For all other substrates, the increase in absorbance at 340 nm ($\varepsilon_{NADH}$ = 6220 $M^{-1}$ $cm^{-1}$), corresponding to the formation of NADH, was monitored. The concentrations of substrate were as follows: **1**, 2-40 μM; **8**, 200-1600 μM, **9**, 200 μM, **10**, 100-1000μM, **11**, 2-50 μM.

To measure the kinetic parameters for $NAD^+$ as substrate, a sufficient quantity of 2-HMSD (120 μL of a 3.0 mg/mL solution) was diluted into buffer (40 mL) and equilibrated at 23 °C for at least 15 min prior to assaying. After transfer of an aliquot (1 mL) to a cuvette, varying amounts of $NAD^+$ (12.5-200 μM) were added by appropriate dilutions of 5, 25, or 50 mM solutions of the coenzyme. Reactions were initiated by the addition of **11** at a saturating condition (200 μM, 10 μL of a 20 mM solution). The increase in absorbance at 340 nm, corresponding to the formation of NADH, was monitored.

*Kinetics of Irreversible Inhibition with Iodoacetamide (**15**).* 2-HMSD (3.0 mg/mL, 58 μM based on the monomer molecular weight) was divided into 100 μL quantities and placed in dark eppendorf tubes at room temperature prior to the addition of inhibitor. Subsequently, varying amounts of **15** (nine concentrations ranging from 75-1000 μM) were added to the incubation mixtures. The volume of inhibitor added did not exceed 10% of the final volume. Aliquots (4 μL) were removed immediately after the addition of inhibitor at 10-s intervals, diluted into 50 mM potassium phosphate buffer (1 mL, pH 8.5) containing 1 mM $NAD^+$, and assayed for residual activity. The assay was initiated by the addition of **11** (200-250 μM). Stock solutions of **15** (50 mM) were made up in 50 mM potassium phosphate buffer (pH 7.4), resulting in a final pH of ~7. The stock solutions were diluted with a sufficient volume of 50 mM potassium phosphate buffer (pH 7.4) to make 0.5 mM and 5 mM solutions. Rates of inactivation ($k_{obs}$) were determined from a least-squares fit of the data obtained from the initial linear portion of the decrease in activity to the equation for a first-order decay using Microsoft Excel. The rate constants ($K_I$ and $k_{inact}$) were determined by fitting the data as previously described (27). Typically, 2-3 experiments were performed at each inhibitor concentration.

*Competitive Inhibition of 2-HMSD.* The reversible competitive inhibition of 2-HMSD was examined using **12**. For these experiments, a quantity of 2-HMSD (95 μL of a 22.8 mg/mL solution) was added to 95 mL of assay buffer [50 mM potassium phosphate buffer (pH 8.5), 1 mM $NAD^+$] and incubated at 22 °C for ~15 min.

Subsequently, the diluted enzyme solution was separated into portions (13 mL) and stored at 22 °C. An aliquot of **12** was added to a portion of the solution to yield the approximate final inhibitor concentration. After ~5 min, aliquots (1 mL) were removed and assayed using eleven concentrations of **1** (2-40 μM). The final concentrations of **12** were 0-100 μM, which were diluted from a stock solution (50 mM) made up in ethanol. The final pH of the enzyme solution was ~8.1 after addition of the inhibitor in all cases. The kinetic data were fitted using the equation for competitive inhibition provided in the Grafit program (Erithacus Software Ltd.).

## RESULTS

*Cloning, Expression, and Purification of C2,3O and 2-HMSD.* The gene for C2,3O has been cloned from a variety of sources. The enzyme from *P. putida* mt-2 has been overexpressed, purified, and crystallized (21, 22). 2-HMSD has previously been purified from *P. putida* mt-2, using either the TOL plasmid pWW0 as the source of the gene or using a constitutively expressed subclone of the TOL plasmid (1, 2). The yield of protein using these sources of protein was not reported. Thus, 2-HMSD has not previously been overexpressed.

In this study, the genes for both C2,3O (*xylE*) and 2-HMSD (*xylG*) were amplified from the TOL plasmid pWW0. C2,3O was required for the synthesis of the compound **1**, the substrate for 2-HMSD. Accordingly, both enzymes were cloned into the T7 expression system for overexpression in *E. coli* strains BL21(DE3)pLysS or BL21-Gold(DE3)pLysS. C2,3O was partially purified (~75% pure as assessed by SDS-PAGE) using a DEAE-5PW anion-exchange column. Further purification was not necessary because the enzyme was used only for the synthesis of **1**.

Purification of 2-HMSD using anion-exchange and hydrophobic interaction chromatography yielded 15-30 mg protein per liter of culture that was essentially (~99%) pure as assessed by SDS-PAGE. DNA sequencing revealed several differences between the published sequence and the sequence from our expression plasmid, pET-24a-2-HMSD, raising a question about the accuracy of either or both sequences. Since the apparent mutations observed in pET-24a-2-HMSD were observed in other clones that had been isolated, and the polymerase used for cloning was a low-error polymerase, the reported sequence is likely not correct. Moreover, the measured kinetic parameters for **1** suggested that the protein generated by this clone was the correct one. Neither the quaternary structure nor the molecular mass of 2-HMSD was determined. The expression systems for both C2,3O and 2-HMSD have yielded high levels of active protein, setting the stage for future studies,

including x-ray crystallography or stopped-flow kinetics, that require large amounts of protein.

*Cloning, Expression, and Purification of the C287A and C287S 2-HMSD Mutants.* Sequence alignment with other aldehyde dehydrogenases (Figure 24) identified several conserved residues that could be involved in the catalytic mechanism of 2-HMSD, including a cysteine, two glutamates, and an asparagine. Cys-287, which aligns with Cys-297 of betaine ALDH, Cys-302 of retinal dehydrogenase, and Cys-243 of rat liver ALDH (Figure 24) was mutated to an alanine and a serine residue to determine whether it was essential for activity. Both mutants were constructed using overlap extension PCR using pET-24a-2-HMSD as template. The resulting plasmids were transformed into BL21(DE3)pLysS for expression and purification. The sequences immediately surrounding the mutations (~400 bp) were confirmed, suggesting that nearby residues have not been changed. Expression of the C287S resulted in insoluble protein. This suggested a disruption in the overall structure of the mutant protein. In light of this finding, further purification of the C287S mutant was not pursued. Expression of the C287A mutant yielded ~50% soluble protein after cell lysis. The soluble protein was purified following the wild-type procedure. The overall yield was 15 mg/L protein.

*Synthesis of 2-HMS.* 2-HMS (**1**) was synthesized using C2,3O, which converts catechol to 2-HMS. It has been previously reported that C2,3O undergoes suicide inactivation by the substrate (25). Thus, pure oxygen was bubbled into the reaction in order to enhance the enzyme's activity. Additionally, the enzyme was added in aliquots over time to reduce the amount of inactivation that might occur. Conversion of catechol to **1** resulted in a decrease in pH, which was adjusted with aliquots of NaOH to maintain neutrality. When the pH no longer decreased, the reaction was considered complete. Extraction and re-crystallization with ethyl acetate resulted in **14** as determined by $^{1}H$ and $^{13}C$ NMR spectroscopy. No further purification was carried out.

*Kinetic Properties of 2-HMSD and C287A-2-HMSD.* In previous studies on 2-HMSD, buffers such as Tris (pH 8.5), glycine-NaOH (pH 9.2-9.6), and CAPS (pH 9.6) were used for enzymatic assays (1, 2, 26). Amine buffers can react with the aldehyde substrates, forming a Schiff base, and thereby affect the results of kinetic studies. Phosphate buffer, however, is inert towards aldehydes. Another reason for investigating the reaction of 2-HMSD in phosphate buffer was to facilitate NMR studies, since the organic buffers would show signals in the NMR spectra, complicating their interpretation. Thus, the activity of 2-HMSD was determined in 50 mM potassium phosphate buffer, pH 8.5, by UV-Vis spectroscopy. The pH was chosen on the basis of published pH rate work, suggesting that 2-HMSD functioned optimally at pH ~8.3 (2).

Initially, 2-HMSD was mistakenly identified as benzaldehyde dehydrogenase from *P. putida* (2, 26). To determine the substrate specificity of the "benzaldehyde dehydrogenase," Harayama and co-workers tested **8** as well as several monosubstituted benzaldehydes with 2-HMSD (2, 26). The substituents included methyl-, methoxy-, halo-, and nitro-groups as well as *p*-carboxybenzaldehyde (**11**). For some of substrates, kinetic parameters were determined. For others, including **11**, only a specific activity in units/mg was described. These determinations were performed at a single substrate concentration, so that no kinetic parameters were reported. Additionally, the initial kinetic parameters for $NAD^+$ were determined using benzaldehyde under conditions that did not saturate the enzyme (10 μM). Later, after correctly identifying the enzyme, 2-HMSD was characterized with **1**, but new parameters for $NAD^+$ were not determined.

In this investigation, **1**, **8**, **11**, and $NAD^+$ were examined, and the previously unstudied carboxy-substituted benzaldehydes (**9**, **10**, Table 1) were also assessed as substrates. Kinetic parameters were obtained for all compounds that showed activity. The kinetic parameters determined among substrates investigated in both this work and in previous work differ to some extent. The kinetic parameters for $NAD^+$ are not comparable because the previous values were not determined under saturating

conditions. Interestingly, the $k_{cat}$ values obtained here for **1** (13.4 $s^{-1}$) and **8** (4.2 $s^{-1}$) are ~10 $s^{-1}$ less than the published values (27.2 and 14.7 $s^{-1}$, respectively). These differences may indicate the non-enzymatic reaction in organic buffers that do not occur in phosphate buffer. The $K_m$ value obtained here for **8** (457 μM) is equal to the reported value (460 μM), while the $K_m$ for **1** (5.7 μM) is three-fold lower than reported (17 μM). Compared to the $k_{cat}/K_m$ value obtained for **1**, **8** shows a 260-fold decrease in activity. The high $K_m$ value observed for **8** is likely due to the absence of a carboxylate, which may be bound by a residue in the active site.

Compounds **9-11** were of interest because they each possess an aldehyde and a carboxylate group. Although **11** was investigated earlier (2), the authors did not recognize the potential similarities between **1** and **11**. Compound **11** is the preferred substrate for 2-HMSD as determined by its $k_{cat}/K_m$ value. The $K_m$ for **1** is 2.6 times greater than that of **11**, while the $k_{cat}$ values are comparable. This results in an overall two-fold increase in $k_{cat}/K_m$. The difference in $K_m$ values may reflect the greater flexibility of **1** in solution, compared to the rigid structure of **11**. Compounds **9** and **10** were poor substrates for 2-HMSD. No activity is observed using **9** as substrate, which may be a result of the adjacent positions of the substituents. Compound **10** is a very poor substrate in comparison to **1** and **11**, with the $k_{cat}/K_m$ values differing 428-fold and 893-fold, respectively. Both $K_m$ (22 to 57-fold) and $k_{cat}$ (16 to 20-fold) also show significant decreases. Although **10** contains both an aldehyde and a carboxylate substituent, the *meta*-positioning of these groups is not the optimal alignment for binding or turnover in the active site.

The kinetic values for $NAD^+$ were not determined using **1** due to the presence of an isobestic point at 340 nm. Compound **11** is the preferred substrate for 2-HMSD as assessed by $k_{cat}/K_m$. Thus, **11** was used at a saturating condition (200 μM, 100x $K_m$) to determine the kinetic parameters for $NAD^+$. The observed $k_{cat}$ is 4 $s^{-1}$, while the $K_m$ is 27 μM. These values cannot be compared to published values because the previous values were determined under non-saturating conditions with a poor substrate (**8**) (1, 26).

C287A 2-HMSD showed no detectable activity by UV-Vis spectroscopy upon incubation with **1** under the conditions used for the wild-type for 30 s. This result was consistent with the sequence analysis implicating Cys-287 in catalysis. The fact that the purified mutant protein eluted under the same conditions as the wild-type suggested that it retained the correct structure despite the mutation. The quaternary structure of the mutant enzyme was not determined.

*Inactivation of 2-HMSD by Iodoacetamide.* Incubation of 2-HMSD with iodoacetamide results in time-dependent inactivation of the enzyme. The decrease in activity is linear for approximately 4 half-lives (Figure 27). The $k_{obs}$ values measured in 26 experiments were plotted versus the inhibitor concentration and fit to a rectangular hyperbola (Figure 28). The values of $k_{inact}$ and $K_I$ obtained from this plot are 0.032 ± 0.007 $s^{-1}$ and 820 ± 290 μM. The hyperbolic inactivation is consistent with the formation of a dissociable complex between the enzyme and iodoacetamide before covalent attachment and inactivation.

*Competitive Inhibition of 2-HMSD by **12**.* On the basis of the structures of the **1** and **2**, and the previous report that 2-HMSD possesses esterase activity (Shaw/Harayama), **12** was examined as a potential inhibitor of 2-HMSD. However, compound **12** did not show evidence of time-dependent inactivation. When an aliquot of 2-HMSD (44.2 μM with respect to the monomer) was incubated with an excess (1 mM) of **12**, the enzyme retained full activity after 13 hr incubation at 4 °C. Subsequently, **12** was examined as a reversible competitive inhibitor. It was found to be a competitive inhibitor with a $K_i$ value of 40 ± 8 μM. This value is 7-fold and 18-fold greater than the $K_m$ values for **1** and **11**, respectively.

## DISCUSSION

Enzymes of the aldehyde dehydrogenase superfamily catalyze the oxidation of an aldehyde as well as the reduction of an acid. While this general reaction has been investigated in many contexts, the ALDH mechanism has not been completely characterized. Human liver aldehyde dehydrogenase is the most notable ALDH, as it is partially responsible for the metabolism of ethanol. Moreover, a mutation in the enzyme gives rise to alcohol intolerance, and drugs such as disulfiram, which are used to treat alcoholics, are known to inhibit this enzyme. A variety of ALDHs have been investigated due to interest in their roles in growth, metabolic and biosynthetic pathways, and detoxification. A sample of the functions that these enzymes serve is demonstrated by the enzymes of Figure 24.

The variety of roles that ALDHs play, as well as the substrates that they accept, suggest reasons for their apparent divergent evolution. Some similarities are shared among members in the superfamily, such as the active-site cysteine. However, differences in structure are to be expected due to differences in substrate and coenzyme specificity. For example, most ALDHs use either $NAD^+$ or $NADP^+$, but not both. It was believed that the binding of the coenzymes would be different in $NAD^+$ or $NADP^+$-dependent enzymes. This was verified by crystallography. In $NAD^+$-dependent ALDHs, a conserved lysine [indicated by a (+) in Figure 24] or a glutamate [indicated by a (¶) in Figure 24] have been reported to interact with the 2′ and 3′ oxygen atoms of the ribose ring of the adenine moiety (8, 10). In $NADP^+$-dependent ALDHs, the conserved lysine interacts with two oxygens of the $NADP^+$ 2′-phosphate. A threonine residue (¶) replaces the glutamate in these enzymes, providing a pocket for the phosphate group of $NADP^+$ (12, 28). Steric hindrance from surrounding residues as well as negative electrostatic interactions are thought to prevent the phosphate group of $NADP^+$ from binding in $NAD^+$-dependent ALDHs (12). Furthermore, differences in accessibility to the active site appear to influence substrate specificity among ALDHs. Enzymes, such as the rat liver ALDH, that

readily oxidize acetaldehyde have channels for diffusion of the substrate. However, retinal dehydrogenase has a disordered substrate channel that may bind long-chain aldehydes using a "hand-in-glove" fit (29).

2-HMSD, which catalyzes the oxidation of 2-HMS (**1**) to the corresponding acid **2**, has not been rigorously examined. The enzyme was initially identified incorrectly as the benzaldehyde dehydrogenase from *P. putida* (2). Although the gene was identified, it was not cloned into a vector for overexpression. The protein was purified to homogeneity using a four-step procedure, and its quaternary structure, corresponding to homodimerization, was determined by gel filtration. Other physical properties were also determined, such as stability, isoelectric point, and stoichiometry with $NAD^+$. Due to its incorrect identification, the enzyme was characterized with benzaldehyde (**8**), mono-substituted analogs possessing methyl-, methoxy-, nitro, chloro-, and fluoro- groups at various positions around the ring, and **11**. For some compounds including **11**, a single substrate concentration was assessed for specific activity, and kinetic parameters were not determined. Kinetic parameters for $NAD^+$ were determined using non-saturating concentrations of a poor substrate (**8**). Assays were performed in organic buffers including Tris, glycine-NaOH, and CAPS, which are amines that could react with the aldehyde moiety. Upon correct identification of the enzyme, the kinetic parameters for **1** were determined, but those for $NAD^+$ were not recalculated using the physiological substrate.

In previous studies, compound **1** was synthesized enzymatically (1), but the procedure used was not reported. A characterization of the compound has not been reported either, apart from the characteristic $\lambda_{max}$ of 375 nm. In addition, **1** is not likely to be stable under aqueous conditions. Compound **1** may undergo ketonization to **5** and **6** in aqueous buffer (Scheme 6). The solution chemistry of **1**, the presumed physiological substrate, has not been characterized.

Mechanistic and structural questions regarding this enzyme have not been addressed. Among these questions are the mechanics of how the reaction is carried out, the identities and functions of crucial residues in catalysis, and the structure of

the enzyme. Although two potential residues in catalysis, Cys-287 and Glu-254, have been identified in previous work, no mutagenesis work was performed (1). In the most recent investigation by Inoue *et al*., crystals of 2-HMSD diffracting to 2.5 Å were obtained (1). However, the corresponding structure has not been published. Identification of the product of the 2-HMSD-catalyzed reaction would suggest the physiological substrate for 4-OT, the next enzyme in the *meta*-fission pathway. The identity of this substrate, either **2** or **3**, has not been directly addressed. Studies to address these many questions require copious amounts of enzyme as well as rigorous characterization of the solution chemistry of a compound (**1**) that is most easily generated by enzymatic synthesis. The results of this investigation represent a significant step forward in addressing these issues and suggest directions for future mechanistic and structural studies.

Our strategy relied on expressing 2-HMSD under control of the T7 promoter, which resulted in large amounts of active protein. 2-HMSD was purified to homogeneity using anion-exchange chromatography followed by hydrophobic interaction chromatography. Due to its apparent highly hydrophobic nature, 2-HMSD eluted from the hydrophobic column during the final buffer wash, resulting in highly concentrated, pure protein. The recombinant protein showed comparable kinetic values to those previously reported for **1**, suggesting that recombinant 2-HMSD is properly folded. Verification of the molecular mass or the quaternary structure of the protein were not performed to further confirm the identity of 2-HMSD due to equipment constraints.

A similar approach was used for the expression of C2,3O, which was required for the synthesis of 2-HMS (**1**), the substrate for 2-HMSD. Although C2,3O was only partially purified (~75%), the expression system generates large amounts of protein, which is useful for the synthesis of 2-HMS. Enzymatic synthesis of 2-HMS yielded only 2-HMS as product. The finding that the isolated compound was **14** was not surprising, since acetals result under acidic conditions and the product is a stable six-membered ring. Compound **14** is stable for at least three years at room temperature.

As a result of the acidification and extraction process, the starting material **13** is readily separated from 2-HMS. Thus, although the yield of 2-HMS (17%) was rather low, **13** could easily be recycled for further synthesis. This property is particularly useful for the generation of isotopically-labelled **1**. Organic synthesis of 2-HMS from commercially available reagents would require considerably more effort and time. C2,3O may be used to oxidize substituted catechols as well, allowing analogs of 2-HMS to be easily synthesized.

We examined several substrates for 2-HMSD in order to probe its structural requirements and to gain insight into potential active site features. Compound **1** is a flexible molecule with many possible conformations, two of which (s-*trans* and s-*cis*) are illustrated in Figure 26. Thus the aldehyde and carboxylate moieties can assume various positions in aqueous solution. To investigate the preferred positioning of these substituents within the enzyme, a series of carboxy-substituted benzaldehydes (**9**-**11**) were tested as substrates. Benzaldehyde (**8**) was also investigated to provide a reference point, apart from **1**, with which the obtained kinetic parameters could be compared.

The results presented in Table 1 indicated that 2-HMSD is able to oxidize compounds **8**, **10**, and **11**. A discrepancy is observed between the published (14.7 $s^{-1}$) and observed (4.2 $s^{-1}$) $k_{cat}$ values for **8**. Since a similar discrepancy of 10 $s^{-1}$ is also observed with **1**, this may be attributable to a non-enzymatic reaction in buffer. However, the observed $K_m$ of 457 μM using **8** agrees with the published $K_m$. It is not surprising that benzaldehyde is a very poor substrate for 2-HMSD, as evidenced by its higher $K_m$. 2-HMS is a long, linear molecule with polar substituents, while benzaldehyde is a cyclic compound that lacks the carboxylate and hydroxyl groups found in 2-HMS. These groups are likely to be important in binding. Indeed, the presence of a carboxylate group in **1** and **11** results in 260-fold and 540-fold rate enhancements, respectively, compared to **8**. This suggests that the carboxylate is important for binding.

The presence of both functional groups alone is not sufficient for catalysis, however. These results also show that the carboxylate group must be properly aligned with respect to the aldehyde. For example, 2-HMSD is not able to process **9**. Although **9** has both functional groups, their *ortho*-positions apparently do not allow **9** to fit correctly into the active site. 2-HMSD processes **10**, in which the two substituents are *meta*- to each other, although poorly. The observed $k_{cat}$ and $k_{cat}/K_m$ values for **10** are the lowest among the investigated substrates. The 3-fold decrease in $K_m$ for **10** compared to **8** again indicates that a carboxylate group improves binding.

Compound **11** is the best substrate for 2-HMSD, as evidenced by its kinetic parameters, probably because its fixed conformation optimally positions the carboxylate and aldehyde groups. In **11**, the carboxylic acid group is maintained *para*- to the aldehyde substituent. This structure mimics one of the possible s-*cis* conformations of **1** in solution (Figure 26). In solution, the s-*trans* conformation is favored due to a decrease in steric hindrance. However, the active site of the enzyme may bind **1** in the s-*cis* conformation that resembles **11** as a result of particular structural residues. The observation that compound **11** is preferred over **1**, the physiological substrate, suggests that the active site may form a channel that preferentially binds molecules in which the carboxylate and aldehyde functional groups are aligned as far apart as possible. A positively charged residue, such as a lysine or an arginine, may reside within the active site, binding the carboxylate group.

The kinetic parameters of 2-HMSD (using **1**) were compared with those of other aldehyde dehydrogenases. Succinic semialdehyde dehydrogenase (SSADH) and aspartate β-semialdehyde dehydrogenase (ASADH) are two other members of the semialdehyde dehydrogenase family (30-32). Although these enzymes prefer different substrates (succinic semialdehyde and $NAD^+$ and β-aspartyl phosphate and $NADP^+$, respectively), general comparisons can be made. For SSADH, the $K_m$ values are 4-15 μM for the aldehyde and 31-130 μM for the nucleotide, and for ASADH, the $K_m$ values are 14 μM for the phosphate and 42 μM for the nucleotide. Thus, these semialdehyde dehydrogenases each have lower $K_m$ values for their aldehyde

substrates compared to the nucleotide coenzymes. Unfortunately, $k_{cat}$ values have not reported, so neither $k_{cat}$ nor $k_{cat}/K_m$ can be compared.

Interestingly, the TOL plasmid of *P. putida* mt-2 encodes a benzaldehyde dehydrogenase. The reported parameters for **8** as a substrate for this enzyme were a $K_m$ of 2.5 μM and a $k_{cat}/K_m$ of 1.7 x $10^7$ (1). These values are similar to the corresponding parameters for 2-HMSD using **1** as a substrate, although there is a 7-fold difference in $k_{cat}/K_m$. Furthermore, although 2-HMSD accepts **8** as a substrate, benzaldehyde dehydrogenase reportedly does not accept **1**. Thus, these enzymes appear to have evolved to fulfill specific roles in the organism which apparently do not overlap in function.

Sequence analysis permitted the identification of putative active site residues. These residues included Asn-155, Cys-287, Glu-254, and Glu-390. Cys-287 is likely the catalytic base as illustrated in Scheme 3. Asn-155 may stabilize the thiohemiacetal intermediate through hydrogen bonding to the oxygen atom. One of the two glutamate residues may deprotonate the cysteine. Crystal structures suggest that the effect of the glutamate may be mediated via a water molecule or a conformational change, since the residue itself seems to be too far (~8 Å) from the aldehyde substrate to directly deprotonate (6, 9).

The role of the cysteine was investigated by affinity labeling and site-directed mutagenesis. Cys-287 was mutated to an alanine and a serine. The C287A mutant retained no apparent activity, which was consistent with a possible role in catalysis. Since the mutant purified in a manner similar to the wild-type, it is likely that it did not suffer from a major change in structure. Additional study is necessary to verify this assumption. It was anticipated that the C287S mutant would exhibit decreased activity due to the presence of a hydroxyl rather than a thiol group. Unfortunately, the C287S mutant resulted in completely insoluble protein, which suggested that the protein was improperly folded. This may have been due to the C287S mutation, or perhaps to the inadvertent mutation of unknown amino acids outside the area that was sequenced were. Further purification and analysis of this mutant was not attempted.

Iodoacetamide is one reagent commonly used in the detection of catalytic cysteines (33). The mechanism of iodoacetamide is illustrated in Scheme 10. In this process, the activated cysteine directly attacks the iodide-substituted methyl group. Iodide is readily displaced because it is a good leaving group. This results in covalent modification of the enzyme, yielding a carboxyamidated cysteine. 2-HMSD was inactivated by iodoacetamide. The inactivation of the enzyme appears to follow saturation kinetics, which is indicative of the formation of a dissociable active site complex prior to inactivation. The precise target for modification was not determined but is currently under investigation.

The results of the site-directed mutagenesis and irreversible inhibition studies clearly implicate C287 in catalysis. Another way to explicitly assign Cys-287 in catalysis is through x-ray crystallography. It would be useful to have either an irreversible or a reversible inhibitor as a ligand in the active site, around which the active site residues could align. Iodoacetamide is not a good potential ligand due to its small size compared to the substrate, **1**, and its potential reactivity over the time required to generate crystals (~2-3 weeks in some cases). Thus the ethyl ester of 2-HM (**12**) was examined as a potential ligand. ALDHs, including 2-HMSD, have been shown to possess esterase activity (2, 34, 35). If 2-HMSD has esterase activity, it might be irreversibly modified (Scheme 11) by **12**. Over time, the enzyme might be re-activated upon exposure to protons in solution. However, time-dependent inactivation was not demonstrated using **12**, indicating that it was not an irreversible inhibitor. However, **12** did prove to be a good competitive inhibitor of 2-HMSD. The $K_i$, although approximately 7 to 18-fold higher than the $K_m$ values for the best substrates, suggests that this compound may be useful as a ligand for structural studies.

UV-Vis and NMR data suggest that the substrate for 2-HMSD is **1** and not **5** or **6**. These data also indicate that the product of 2-HMSD is **2**. Thus, 4-OT, the next enzyme in the pathway, is required for the processing of **2** to afford **4**, which is processed by 4-OD as substrate. These data may be further verified by obtaining the

structure of 2-HMSD with **12**. It would be interesting to determine whether or not **1** and/or **2** tautomerize within the active site of 2-HMSD.

The results presented herein set the stage for further analysis of 2-HMSD. A number of experiments can still be done to delineate the roles of active site residues. The three other putative active site residues identified by sequence alignment (Asn-155, Glu-254, and Glu-390) were not analyzed in this work, but they could be mutated to determine the effects on the activity of the protein. An x-ray structure of 2-HMSD can be pursued because both an efficient expression and purification procedure have been developed. The characterization of a potent reversible competitive inhibitor, **12**, makes a ligand for such studies available. Stopped-flow or rapid quench experiments may be used to "trap" the thiohemiacetal intermediate, which has been observed. Future studies of 2-HMSD will not only shed light on its mechanism, but also provide insight as to this enzyme's evolutionary relationship to the other members of the ubiquitous aldehyde dehydrogenase superfamily.

**ACKNOWLEDGMENTS**

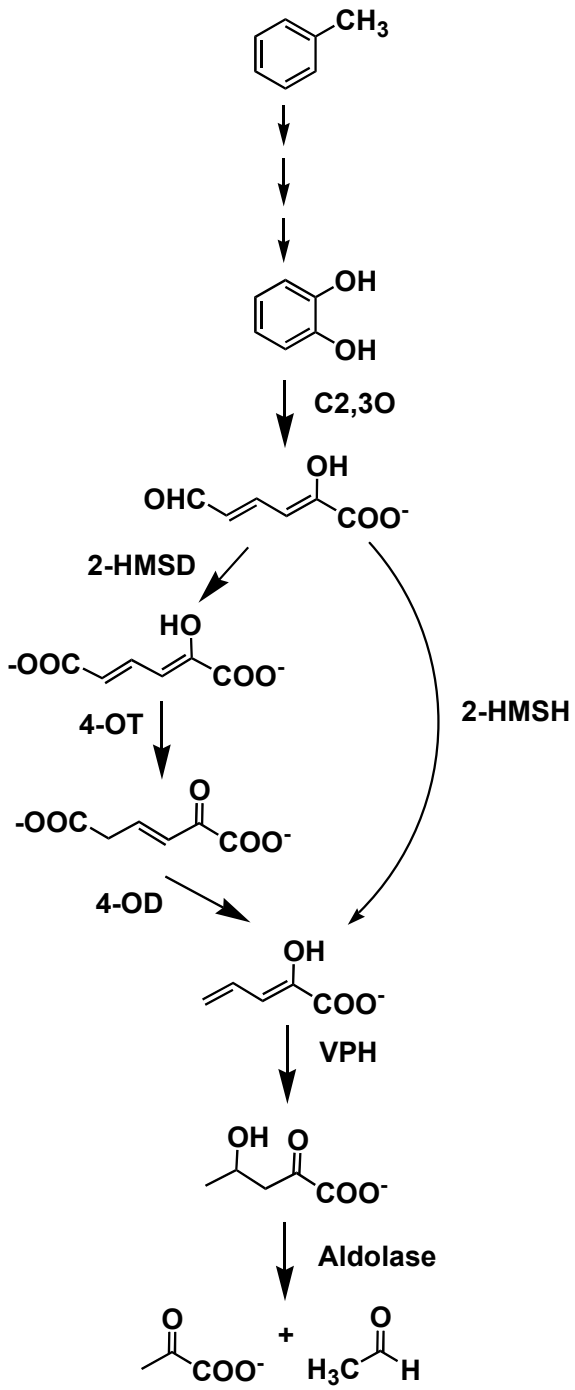
We thank Steve D. Sorey (Department of Chemistry, The University of Texas at Austin) for his expert assistance in acquiring the NMR spectra, and Dr. William H. Johnson, Jr. (Division of Medicinal Chemistry, The University of Texas at Austin) for synthesizing needed compounds and interpreting the NMR spectra.

**REFERENCES**

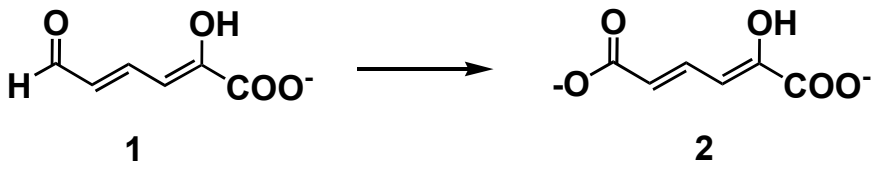

1. Inoue, J., Shaw, J. P., Rekik, M., and Harayama, S. *J. Bact.*, 1995, **177**, 1196-1201.
2. Shaw, J. P., and Harayama, S. *Eur. J. Biochem.*, 1990, **191**, 705-714.
3. Vasiliou, V., Pappa, A., and Petersen, D. R. *Chem. Biol. Interact.*, 2000, **129**, 1-19.
4. Lundblad, R. L. and Noyes, C. M. Chemical Reagents for Protein Modification, 1984.
5. Zhang, L, Ahvazi, B., Szittner, R., Vrielink, A., and Meighen, E. *Biochemistry,* 2000, **39**, 14409-14418.
6. Hurley, T. D., Steinmetz, C. G., and Weiner, H. *Adv. Exp. Med. Biol.*, 1999, **463**, 15-25.
7. Wang, X., and Weiner, H., *Biochemistry*, 1997, **34**, 237-243.
8. Liu, Z.-J., *et al. Nature Struc. Biol.*, 1997, **4**, 317-326.
9. Hempel, J., *et al. Adv. Exp. Med Biol.*, 1999, **463**, 53-59.
10. Johansson, K., *et al. Prot. Sci.,* 1998, **7**, 2106-2117.
11. Pietruszko, R., Kikonyogo, A., Chern, M.K., and Izaguirre, G. *Adv. Exp. Med. Biol.,* 1997, **414**, 243-252.
12. Cobessi, D., *et al.*, *J. Mol Biol.*, 1999, **290**, 161-173.
13. Bognar, A., and Meighen, E. A., *J. Biol. Chem.*, 1978, **253**, 446-450.
14. Rossmann, M. G., Moras, D., and Olsen, K. W. *Nature*, 1974, **250**, 194-199.
15. Blanco, J., Moore, R. A., Kabaleeswaran, V., and Viola, R. E., *Protein Science*, 2003, **12**, 27-33.
16. Whitman, C. P., Aird, B. A., Gillespie, W. R., and Stolowich, N. J., *JACS,* 1991, **113**, 3154–3162.
17. Stanley, T. M., *et al. Biochemistry*, 2000, **39**, 718-726.
18. Sambrook, J., Fritsch, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, 1989.

19. Laemmli, U. K. *Nature*, 1970, **227**, 680-685.
20. Waddell, W. J. *J. Lab. Clin. Med.,* 1956, **48**, 311-314.
21. Kobayashi, T., *et al. J. Biochem. (Tokyo)*, 1995, **117**, 614-622.
22. Kita, A., *et al. Structure Fold Des.*, 1999, **7**, 25-34.
23. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. *Gene,* 1989, **77**, 51-59.
24. Nakai, C., *et al. J. Biol. Chem.*, 1983, **258**, 2923-2928.
25. Cerdan, P., Wasserfallen, A., Rekik, M., Timmis, K. N., and Harayama, S. *J. Bact.*, 1994, **176**, 6074-6081
26. Shaw, J. P., Schwager, F., and Harayama, S. *Biochem. J.*, 1992, **283**, 789-794.
27. Stivers, J. T., *et al*. *Biochemistry*, 1996, **35**, 803-813.
28. Ahvazi, B., *et al. Biochem. J.*, 2000, **349**, 853-861.
29. Lamb, A. L. and Newcomer, M. E. *Biochemistry*, 1999, **38**, 6003-6011.
30. Karsten, W. E., and Viola, R.E. *Biochim. Biophys. Acta*, 1991, **1077**, 209-219.
31. Busch, K. B., and Fromm, H. *Plant Phys.*, 1999, **121**, 589-597.
32. Satya Narayan, V., and Nair, P. M. *Arch. Biochem. Biophys.,* 1989, **275**, 469-477.
33. Means, G. E., and Feeney, R. E. Chemical Modification of Proteins, 1971.
34. Ting, H. H., and Crabbe, M. J., *Biochem. J.*, 1983, **215**, 361-368.
35. Fong, W., and Choi, K., *Chem. Biol. Interact.*, 2001, **130-132**, 161-171.

**Scheme 1**

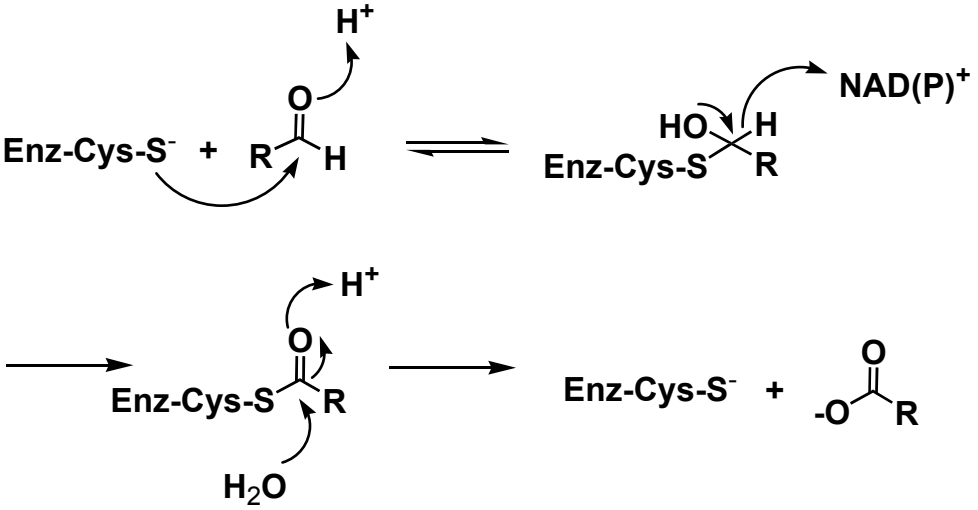

CH3
OH
OH
C2,3O
OH
OHC
COO-
2-HMSD
HO
-OOC
COO-
4-OT
O
-OOC
COO-
4-OD
2-HMSH
OH
COO-
VPH
OH
O
COO-
Aldolase
O
COO-
+
O
H3C
H

**Scheme 2**

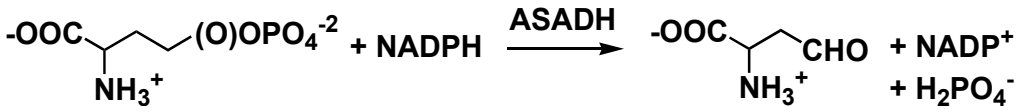



1 2

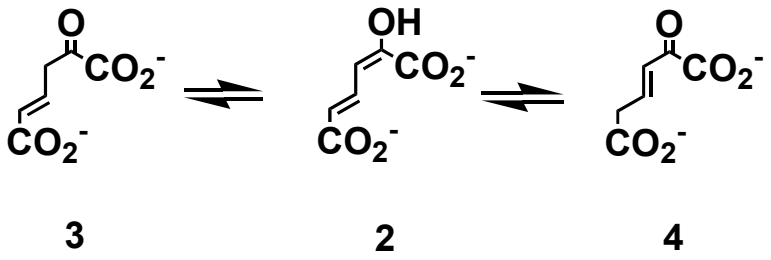
**Scheme 3**



**Scheme 4**

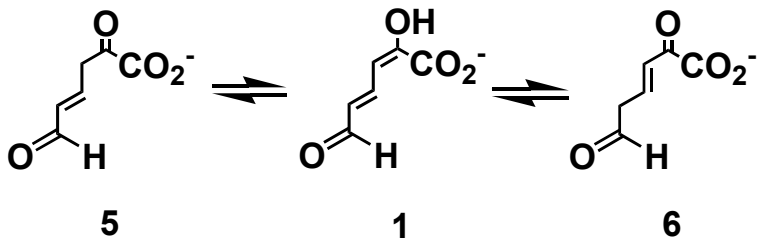


**Scheme 5**

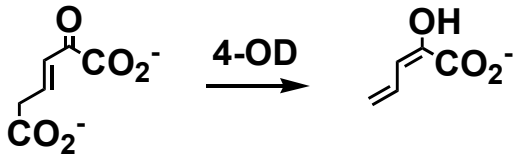



**3** **2** **4**

**Scheme 6**

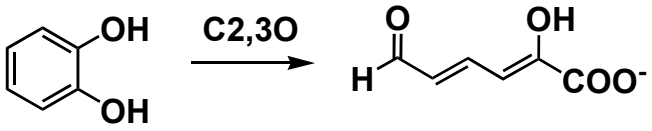


**Scheme 7**



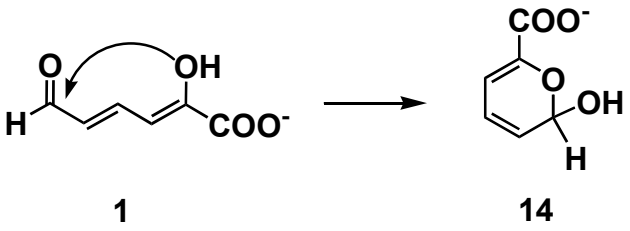

**4** **7**

**Scheme 8**



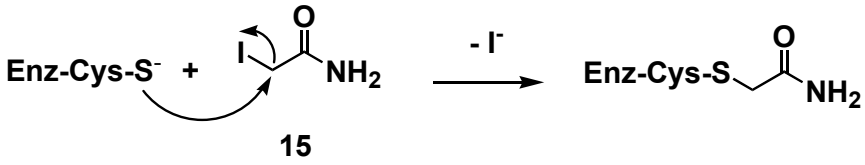

**13** **1**

**Scheme 9**




**1** **14**

**Scheme 10**



**Scheme 11**

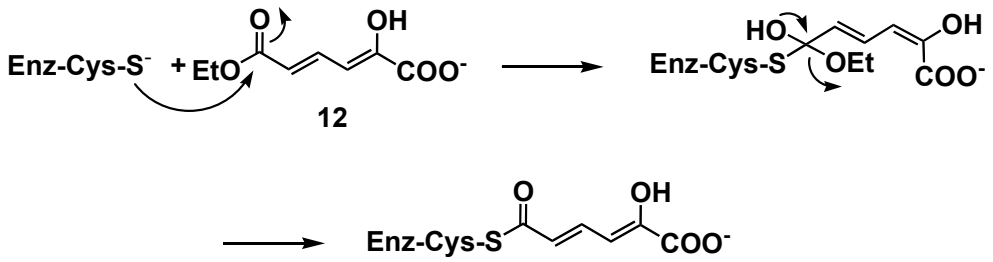


**Figure 24.** Sequence alignment of aldehyde dehydrogenases (ALDHs) generated using CLUSTAL-W. From top to bottom, betaine dehydrogenase, retinal dehydrogenase, 2-hydroxymuconate semialdehyde dehydrogenase, *Streptococcus mutans* ALDH, rat ALDH, *Vibrio harveyi* ALDH.

(*) indicates conserved residues. (-) indicates the Rossman fold. (+) and (¶) indicate residues for nucleotide binding. (‡) indicates an unconserved histidine implicated in catalysis. Symbols are located above the appropriate residue(s).

```
betaine   AQLVDSMPSASTGSV VVTDDLNYWGGRRIK SKDGATTEPVFEPAT GRVLCQMVPCGAEEV DQAVQSAQAAYLK-- -WSKMAGIERSRVML     87
retinal   -ASLQLLPSPTPN-- LEIKYTKIFINNEWQ NSESGRVFPVCNPAT GEQVCEVQEADKVDI DKAVQAARLAFSLGS VWRRMDASERGRLLD     88
hmsd      --------------- ---KEIKHFISGELV GSASGKLFDNVSPAN GQVIGRVHEAGRAEV DAAVRAARAALKG-- PWGKMTVAERAEILH     71
Strep     --------------- --TKQYKNYVNGEWK LSENE--IKIYEPAS GAELGSVPAMSTEEV DYVYASAKKAQPA-- -WRALSYIERAAYLH     69
rat       --------------- --------------- --------------- ------------SSI SDTVKRAREAFNS-- -GKTRSLQFRIQQLE     31
vibrio    --------------- NPQTDNVFYATNAFT GEALPLAFPVHTEVE VNQAATAAAKVARDF RRLNNSKRASLLR-- ------TIAS-ELEA     67
                                                                                                  β1
                                                                                                  ----    * ---
betaine   EAARIIRERRDNIAK LEVINNGK-TITEAE YDIDAAWQCIEYYAG LAPTLSGQHIQLPG- -----GAFAYTRREP LGVCAGILAWNYPFM    170
retinal   KLADLVERDRATLAT MESLNGGKPFLQAFY IDLQGVIKTLRYYAG WADKIHGMTIPVDG- -----DYFTFTRHEP IGVCGQIIPWNFPLL    172
hmsd      RVADGITARFGEFLE ARMPGHRQAEVAGQP HRHSARRANFKVFAD LLKNVANEAFEMATP DGA--GALNYGVRRP KGVIGVISPWNLPLL    159
Strep     KVADILMRDKEKIGA ILSKEVAK-GYKSAV SEVVRTAEIINYAAE EGLRMEGEVLEGGSF EAASKKKIAVVRREP VGLVLAISPFNYPVN    158
rat       ALQRMINENLKSISG ALASDLGKNEWTSYY EEVAHVLEELDTTIK ELPDWAEDEPVAKTR QTQ--QDDLYIHSEP LGVVLVIGAWNYPFN    119
vibrio    RSDDIIARAHLETAL PEVRLTGEIARTANQ LRLFADVVNSGSYHQ AILDTPNPTRAPLP- -----KPDIRRQQIA LGPVAVFGASNFPLA    151
             a1             β2            α2                 β3          α3          β4          α4
          ----------        ---+  ¶  ----------------          -----       -------      -----    ---------
betaine   IAAWK--CAPALACG NAVVFKPSPMTPVTG VILAEIFHEAG---- VPVGLVNVVQG--GA ETGSLLCHHPNVAKV SFTGSVPTGKKVMEM    252
retinal   MFTWK--IAPALCCG NTVVIKPAEQTPLSA LYMGALIKEAG---- FPPGVVNILPGY-GP TAGAAIASHIGIDKI AFTGSTEVGKLIQEA    255
hmsd      LMTWK--VGPALACG NCVVVKPSEETPLTA TLLGEVMQAAG---- VPAGVYNVVHGFGGD SAGAFLTEHPDVDAY TFTGETGTGETIMRA    243
Strep     LAGSK--IAPALIAG NVIAFKPPTQGSISG LLLAEAFAEAG---- LPAGVFNTITGR-GS EIGDYIVEHQAVNFI NFTGSTGIGERIGKM    241
rat       LTIQP--MVGAVAAG NAVILKPSEVSGHMA DLLATLIPQYM---- DQNLYLVVKGG---- VPETTELLKERFDHI MYTGSTAVGKIVMAA    199
vibrio    FSAAGGDTASALAAG CPVIVKGHTAHPGTS QIVAECIEQALKQEQ LPQAIFTLLQGN-QR ALGQALVSHPEIKAV GFTGSVGGGRALFNL    240
                  β5
          ---     ---*-
betaine   SAKT--VKHVTLELG GKSPLLIFKDCELEN AVRGALMANFLTQ-- GQVCTNGTRVFVQRE IMPQFLEEVVKRTKA IVVGDPLLT----ET    334
retinal   AGRSN-LKRVTLELG GKSPNIIFADADLDY AVEQAHQGVFFNQ-- GQCCTAGSRIFVEES IYEEFVKRSVERAKR RIVGSPFDP----TT    338
hmsd      AAKG--VRQVSLELG GKNAGIVFADCDMDK AIEGTLRSAFANC-- GQVCLGTERVYVERP IFDAFVARLKAGAEA LKIGEPNDP----EA    325
Strep     AGMR----PIMLELG GKDSAIVLEDADLEL TAKNIIAGAFGYS-- GQRCTAVKRVLVMES VADELVEKIREKVLA LTIGNPED-----DA    320
rat       AAKH--LTPVTLELG GKSPCYVDKDCDLDV ACRRIAWGKFMNS-- GQTCVAPDYILCDPS IQNQIVEKLKKSLKD FYGEDAKQ-----SR    280
vibrio    AHERPEPIPFYGELG AINPTFIFPSAMRAK ADLADQFVASMTMGC GQFCTKPGVVFALNT PETQAFIETAQSLIR QQSPSTLLTPGIRDS    330
                                                                                          *
betaine   RMGGLISKPQLDKVL GFVAQAKK-----EG ARVLCGGEPLTPSDP KLKNGYFMSPCVLDN CRDDMTCVKEEIFGP VMSVLPFDTEEEVLQ    419
retinal   EQGPQIDKKQYNKIL ELIQSGVA-----EG AKLECGGKGLG---- --RKGFFIEPTVFSN VTDDMRIAKEEIFGP VQEILRFKTMDEVIE    417
hmsd      NFGPLISHKPREKVP SYYQQAVD-----DG ATVVTGGGVPEMP-A HLAGGAWVQPTIWTG LADDSAVVTEEIFGP CCHIRPFDSEEEAIE    409
Strep     DITPLIDTKSADYVE GLINDAND-----KG ATALTEIKREG---- -----NLICPILFDK VTTDMRLAWEEPFGP VLPIIRVTSVEEAIE    396
rat       DYGRIINDRHFQRVK GLIDNQ--------- -KVAHGGTWDQ---- ---SSRYIAPTILVD VDPQSPVMQEEIFGP VMPIVCVRSLEEAIQ    353
vibrio    YQSQVVSRGSDDGID VTFSQAESPCVASAL FVTSSENWRKHPAWE EEIFGPQSLIVVCEN VADMLSLSEMEAGSL TATIHATEEDYPQVS    420
                                  ‡
betaine   RANNTTFGLASGVFT RDISRAHRVAANLEA GTCYINTYSIS--PV EVPFGGYKMSGFGRE NGQATVDYYSQLKTV IVEMGDVDSLF----    503
retinal   RANNSDFGLVAAVFT NDINKALMVSSAMQA GTVWINCYNAL--NA QSPFGGFKMSGNGRE MGEFGLREYSEVKTV TVKIPQKNS------    499
hmsd      LANSLPYGLASAIWT ENVRRAHRVAGQIEA GIVWVNSWFLR--DL RTAFGGSKQSGIGRE GGVHSLEFYTELKNI CVKL-----------    486
Strep     ISNKSEYGLQASIFT NDFPRAFGIAEQLEV GTVHINNKTQRG-TD NFPFLGAKKSGAGIQ GVKYSIEAMTTVKSV VFDIK----------    475
rat       FINQREKPLALYVFS NNEKVIKKMIAETSS GGVTANDVIVHITVP TLPFGGVGNSGMGAY HGKKSFETFSHRRSC LVKSL----------    433
vibrio    QLIPRLEEIAGRLVF NGWPTGVEVGYAMVH GGPYPASTHSASTSV GAEAIHRWLRPVAYQ ALPESLLPDSLKAEN PLEIARAVDGKAAHS    510
```

**Figure 25**. Top, potential substrates for 2-HMSD: **8**, benzaldehyde, **9**, 2-carboxylbenzaldehyde, **10**, 3-carboxybenzaldehyde, **11**, 4-carboxybenzaldehyde. Bottom, potential inhibitor of 2-HMSD.

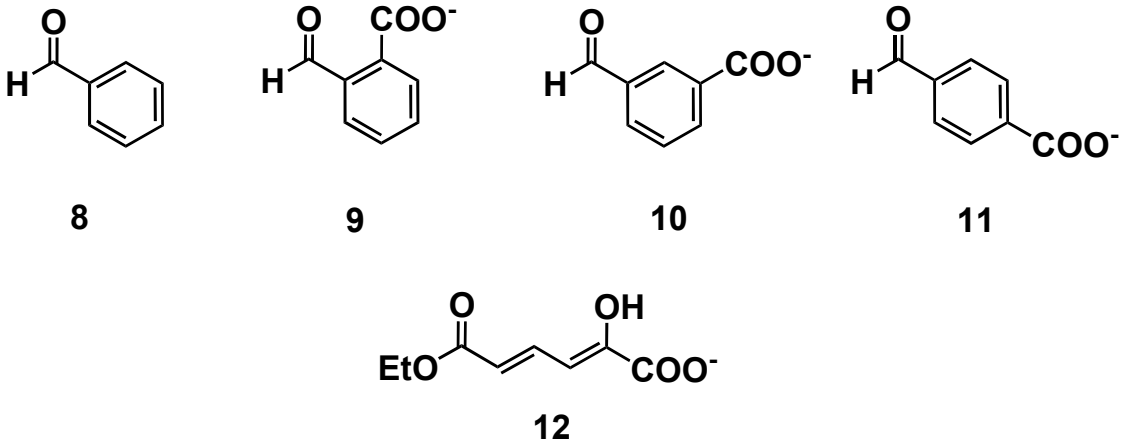
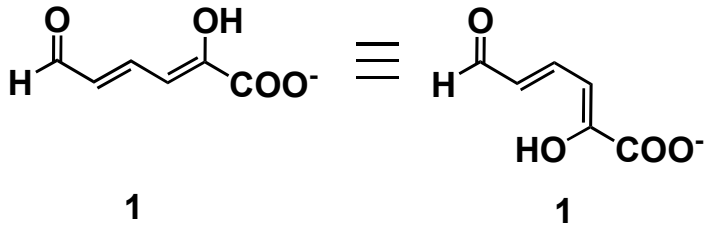
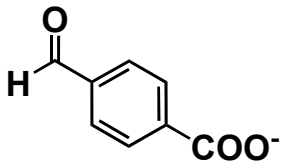


**Figure 26**. Top, two possible conformations of 2-HMS (**1**). Left, s-*trans*, right, s-*cis*. Bottom, 4-carboxybenzaldehyde (**11**).




**1** **1**




**11**

**Figure 27**. The inactivation of 2-HMSD after incubation with varying amounts of **15** (filled triangles, 75 μM; empty triangles, 100 μM; empty diamonds, 150 μM; filled diamonds, 250 μM; empty squares, 400 μM; filled squares, 800 μM; empty circles, 1000 μM). The concentration of 2-HMSD was 58 μM.

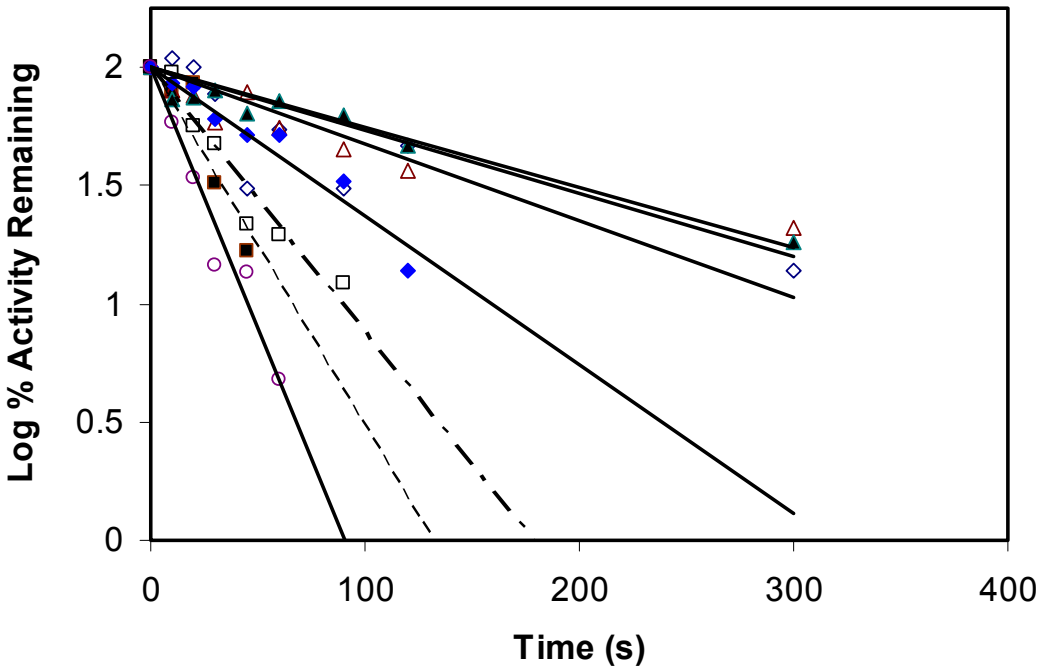


**Figure 28**. Determination of the $k_{inact}$ and $K_I$ values for **15**. A plot of the $k_{obsd}$ values for inactivation measured in 26 experiments as a function of varying amounts of 7 (75-1000 μM). The values of $k_{inact}$ and $K_I$ obtained from this plot are 0.032 ± 0.007 $s^{-1}$ and 820 ± 290 μM, respectively.

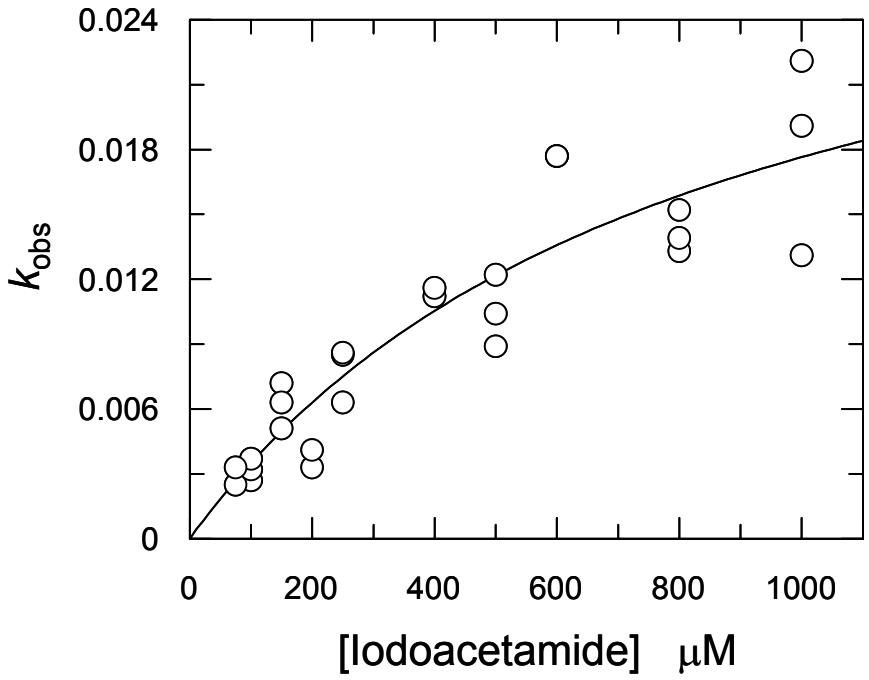


**Table 1. Kinetic Parameters for HMSD[a]**

| substrate | $k_{cat}$ ($s^{-1}$) | $K_m$ ($\mu M$) | $k_{cat}/K_m$ ($M^{-1}\ s^{-1}$) |
|---|---|---|---|
| **1** | 13.4 ± 1.1 | 5.7 ± 1.1 | $2.4 \times 10^6$ |
| **8** | 4.2 ± 0.5 | 457.2 ± 142.6 | $9.2 \times 10^3$ |
| **9** | No activity | No activity | No activity |
| **10** | 0.7 ± 0.02 | 124.3 ± 13.9 | $5.6 \times 10^3$ |
| **11** | 11.1 ± 0.4 | 2.2 ± 0.4 | $5.0 \times 10^6$ |
| $NAD^+$ | 4.2 ± 0.2 | 27.2 ± 4.2 | $1.5 \times 10^5$ |

[a]The steady-state kinetic parameters were determined in 50 mM potassium phosphate buffer (pH 8.5) at 23 °C. Errors are standard deviations**.**

# REACTIONS OF *TRANS*-3-CHLOROACRYLIC ACID DEHALOGENASE WITH ACETYLENE SUBSTRATES: CONSEQUENCES OF AND EVIDENCE FOR A HYDRATION REACTION

## INTRODUCTION

The enzyme *trans*-3-chloroacrylic acid dehalogenase (CaaD) converts the *trans*-isomers of 3-bromo- and 3-chloroacrylates (**1**, Scheme 12) to malonate semialdehyde (**3**) (1-3). CaaD has been isolated from *Pseudomonas pavonaceae* strain 170 and is part of a pathway that enables the bacterium to use either 3-chloroacrylate or 1,3-dichloropropene as sources of carbon and energy (4). The latter compound is one of the active ingredients in the commercial products Shell D-D and Telone II, which are mixed in with the soil and used to control plant-parasitic nematodes (5).

The genes for the dehalogenase were recently cloned from the *P. pavonaceae* strain, the gene expressed, and the enzyme purified and characterized (3). The enzyme is a heterohexamer, consisting of 3 α-subunits, each having 75 amino acid residues, and 3 β-subunits, each having 70 amino acid residues. Sequence analysis revealed similarities between the α-subunit of CaaD and a bacterial isomerase, 4-oxalocrotonate tautomerase (4-OT), and the β-subunit of CaaD and a 4-OT homologue found in *Bacillus subtilis* and designated YwhB. Both 4-OT and YwhB are homohexamers, with monomers that consist of 62 and 61 amino acids, respectively (6, 7).

The mechanism of 4-OT has been studied extensively over the past decade (7). The amino-terminal proline functions as a general base, converting 2-oxo-4-hexenedioate (**4**, Scheme 13) to 2-oxo-3-hexenedioate (**6**) through a dienol intermediate, known commonly as 2-hydroxymuconate (**5**) (7). Arg-39 and an ordered water molecule provide hydrogen bonds to the carbonyl oxygen of **4**, thereby facilitating deprotonation at C-3 (8, 9). Arg-11 interacts with the C-6 carboxylate group and functions as an electron sink to draw electron density to the C-5 position

for protonation. Phe-50 plays a role in maintaining the hydrophobicity of the active site, which assists in catalysis and contributes to the lowered p$K_a$ of Pro-1 (10). Significantly, both the α– and β– subunits of CaaD have an amino-terminal proline while the α-subunit has Arg-11 (3).

These similarities and the results of mutagenesis experiments prompted Janssen and co-workers to suggest a mechanism involving the addition of water to the double bond of **1** to form the unstable chlorohydrin intermediate, **2** (3). Subsequent collapse to **3** can occur by either the direct expulsion of the chloride or by the α,β-elimination of HCl, followed by ketonization (11). Pro-1 of the β-subunit was identified as a critical mechanistic residue because the P1A mutant had no activity while mutation of Pro-1 in the α-subunit (to an alanine) had little effect on activity (3). Two possible roles were proposed for Pro-1 (3). It could function as a general base to activate the water molecule for attack at C-3 of **1** (as shown in Scheme 14) or it could function as a general acid to protonate C-2 of **1** while the water molecule is activated in some other manner. Arg-11 of the α-subunit was also identified as a critical residue because the R11A mutant was inactive. It is proposed that Arg-11 interacts with the C-1 carboxylate group of the 3-haloacrylate.

The mechanism of CaaD and the parallels between it and 4-OT are intriguing and raise several questions. One of the more exciting questions concerns how the two enzymes evolved and diverged to catalyze such different reactions. In order to gain a better mechanistic understanding of CaaD, and to ultimately address this evolutionary question, we conducted a kinetic analysis of CaaD and examined its behavior with three acetylene substrates. In the course of this investigation, we developed an efficient expression system for CaaD as well as a direct UV spectrophotometric assay to monitor its activity, and verified that **3** is the product of the reaction. We further determined that CaaD converts 2-oxo-3-pentynoate (**7**, Scheme 15) to acetopyruvate (**8**), suggesting that 3-bromo- and 3-chloropropiolioate (**9** and **10**, respectively) might function as mechanism-based inhibitors of CaaD. Both compounds were found to be potent irreversible inhibitors of CaaD with the sole target of modification by **9** being

Pro-1 in the β-subunit. Thus, CaaD hydrated the three acetylene compounds, but the resulting products had different consequences for the enzyme. These observations are consistent with an enzyme-catalyzed hydration reaction and indicate that the 3-halopropiolic acids are promising candidates for further mechanistic and structural studies.

## MATERIALS AND METHODS

*Materials.* All reagents, buffers, and solvents were obtained from Aldrich Chemical Co. (Milwaukee, WI), Sigma Chemical Co. (St. Louis, MO), Fisher Scientific Inc. (Pittsburgh, PA), Spectrum Laboratory Products, Inc (New Brunswick, NJ), or EM Science (Cincinnati, OH) with the following exceptions. The *trans*-3-chloroacrylate was obtained from Fluka Chemical Corp. (Milwaukee, WI). Literature procedures were used for the syntheses of 2-oxo-3-pentynoate (**7**) (12), 3-bromo- and 3-chloropropiolic acids (**9** and **10**), and the *trans*-isomer of 3-bromoacrylate (**11**) (13, 14). Tryptone, yeast extract, and agar were obtained from Becton, Dickerson, and Company (Franklin Lakes, NJ). The Ultrafree DA centrifugal filter units, the YM-3 ultrafiltration membranes, and the Amicon concentrators were obtained from Millipore Corp. (Billerica, MA). The thin-walled PCR tubes were obtained from Ambion, Inc. (Austin, TX). Restriction enzymes, PCR reagents, and T4 DNA ligase were obtained from F. Hoffmann-La Roche, Ltd. (Basel, Switzerland), Promega Corp. (Madison, WI), or New England Biolabs (Beverly, MA). The Wizard PCR Preps DNA purification system was purchased from Promega Corp. The QIAprep Spin Miniprep kit was obtained from Qiagen, Inc. (Valencia, CA). The remaining molecular biology reagents including agarose, DNA ladders, and protein molecular weight standards, were obtained from Invitrogen Corporation (Carlsbad, CA). Oligonucleotides for DNA amplification and sequencing were synthesized by either Oligos Etc. (Wilsonville, OR) or Genosys (The Woodlands, TX).

*Strains. Escherichia coli* strains DH5α (Invitrogen) and JM109 (Promega) were used for cloning and isolation of plasmids. *E. coli* strains BL21(DE3)pLysS from Novagen and BL21-Gold(DE3)pLysS from Stratagene (La Jolla, CA) were used for recombinant protein expression.

*Methods.* Techniques for restriction enzyme digestions, ligation, transformation, and other standard molecular biology manipulations were based on methods described elsewhere (15), or as suggested by the manufacturer. Plasmid DNA was introduced into cells by electroporation using a Cell-Porator

Electroporation System (Whatman Biometra, Göttingen, Germany). The PCR was performed in a DNA thermal cycler (Model 480) obtained from PerkinElmer Inc (Wellesley, MA). Colony screening by the PCR is based on methods described in the pET System Manual (Novagen). DNA sequencing was done at the DNA Core Facility in the Institute for Cellular and Molecular Biology at the University of Texas (Austin). Kinetic data were obtained on a Hewlett Packard 8452A or an Agilent 8453 Diode Array spectrophotometer. The cuvettes were mixed using a stir/add cuvette mixer (Bel-Art Products, Pequannock, NJ). The kinetic data were fitted by nonlinear regression data analysis using the Grafit program (Erithacus Software Ltd., Horley, U. K.) obtained from Sigma Chemical Co. HPLC was performed on a Waters (Milford, MA) 501/510 system or a Beckman System Gold HPLC (Fullerton, CA) using a TSKgel Phenyl-5PW hydrophobic column (TosoHaas, Montgomeryville, PA) or a Superose 12 gel filtration column (Amersham Biosciences, Piscataway, NJ). Protein was analyzed by Tris glycine sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions on 17.5% gels using either the Mini-Protean II or III vertical gel electrophoresis apparatus obtained from Bio-Rad (Hercules, CA) (16). Protein concentrations were determined using the method of Waddell (17). The NMR spectra were recorded in 100% $H_2O$ on a Varian Unity INOVA-500 spectrometer using selective pre-saturation of the water signal with a 2-s pre-saturation interval. The lock signal is dimethyl-$d_6$ sulfoxide. Chemical shifts are standardized to the dimethyl-$d_6$ sulfoxide signal at 2.49 ppm.

*Synthesis of the α- and β-Subunit Genes and Construction of the Expression Vectors*. In separate reactions, the genes for the α- and β-subunit of CaaD were synthesized in two stages in a Perkin Elmer DNA Thermocycler using either five (for the α-subunit) or four (for the β-subunit) long oligonucleotides, two external oligonucleotide primers, and the PCR reagents (18-20). The five long oligonucleotides (designated α1-α5) used for the synthesis of the α-subunit along with the two external oligonucleotide primers (designated primers A and B) used for amplification of the synthesized gene are listed in Table S1. Primer A consists of

nine bases to facilitate digestion by *Nde*I, the *Nde*I restriction site (underlined), and 15 bases corresponding to the coding sequence of the first five amino acids of the α-subunit. Primer B consists of nine bases to facilitate digestion by *Not*I, the *Not*I restriction site (underlined), the complementary sequence for a stop codon, and 18 bases corresponding to the complementary sequence of the last six amino acids of the α-subunit. Oligonucleotides α1, α3, and α5 represent the coding base sequences for 17-, 24-, and 14-amino acid segments, respectively, of the α-subunit. These segments correspond to the initiating methionine (designated as the first amino acid) to Arg-18, Ala-31 to His-54, and Pro-63 to the stop codon, respectively. Oligonucleotides α2 and α4 represent the complementary base sequences coding for 24- and 20-amino acid segments, respectively, of the α-subunit. These segments correspond to Thr-13 to Arg-36 and Ile-49 to Asp-68, respectively. Adjacent oligonucleotides overlap by 18 bases. The first 18 bases of oligonucleotide α1 are identical to the last 18 bases of primer A, while the last 18 bases of oligonucleotide α5 correspond to the complementary sequence for the last 18 bases of primer B.

The four long oligonucleotides (designated β1-β4) used for the synthesis of the β-subunit along with the two external oligonucleotide primers (designated primers C and D) used for amplification of the synthesized gene are listed in Table S2. Primer C was designed so that the gene for the β-subunit could either be co-expressed with the α–subunit on the same plasmid or expressed separately from its own plasmid. Thus, it contains an initial eight bases (to facilitate digestion) followed by sites for both the *Not*I and *Nde*I (underlined) restriction enzymes. Separating the two sites are a stop codon, the sequence for a ribosomal binding site (GAAGGA), and six bases (GATATA), which correspond to the sequence found in the pET-21a(+) vector. Primer D contains a *Not*I restriction site (underlined) followed by the complementary sequence for a stop codon and 17 bases corresponding to the complementary sequence of the last six amino acids of the β-subunit. Oligonucleotides β1 and β3 represent the coding sequences for 18- and 24-amino acid segments, respectively, and correspond to Met-1 to Gln-18 and Ser-31 to Ser-54. Oligonucleotides β2 and β4

represent the complementary base sequences for 24- and 23-amino acid segments, which correspond to Ser-13 to Pro-36 and Ala-49 to the stop codon, respectively. The only bases shared by both primer C and oligonucleotide β1 are those corresponding to the start codon (i.e., ATG). The last 18 bases of primer D are the same as the first 18 bases of oligonucleotide β4. Again, adjacent oligonucleotides overlap by 18 bases.

In stage 1 of the PCR for the synthesis of the α-subunit, a 50-μL reaction mixture consisted of the 10× PCR reaction buffer (5 μL), the five long oligonucleotides (2 μl each from 1 μM solutions), primers A and B (2.5 μL each from 20 μM solutions), dNTPs (1 μL from a 10 mM stock solution containing the four dNTPs), 25mM $MgCl_2$ (3 μL), Taq polymerase (1 μL), and sterile de-ionized water (26 μL). The reaction mixture was contained in a thin-walled reaction tube and layered with sterile paraffin oil. The PCR protocol was made up of an initial 5-min denaturation cycle at 94 °C, 25 cycles of PCR, and a final 5-min elongation period at 72 °C. Each cycle consisted of three steps: denaturation at 94 °C for 1 min, annealing at 55 °C for 75 s, and elongation at 72 °C for 2 min. The amplified PCR products were analyzed by electrophoresis on 1% agarose gels. Two distinct bands were observed, one at approximately 200 bp and one at approximately 300 bp. The bands were separately excised and purified using the Wizard PCR Preps DNA purification system.

In the second stage of PCR, these purified PCR products were used as templates in individual PCRs. Accordingly, each reaction contained 10× PCR reaction buffer (5 μL), a purified PCR product (1 μL), primers A and B (2.5 μL each from 20 μM solutions), dNTPs (1 μL from a 10 mM stock solution containing the four dNTPs), 25mM $MgCl_2$ (3 μL), Taq polymerase (1 μL), and sterile de-ionized water (34 μL). After subjecting the reaction mixtures to the PCR protocol described above, the amplified PCR products were analyzed by electrophoresis on 1% agarose gels. Both PCRs gave rise to similar sets of bands. The most intense band, observed at

approximately 250 bp, was excised and purified using an Ultrafree-DA centrifugal unit.

Subsequently, the purified PCR product and the pET-24a(+) vector were treated with the *Nde*I and *Not*I restriction enzymes overnight at 37 °C, and the digestion products were electrophoresed on a 1% agarose gel. Further purification was achieved by use of the Ultrafree-DA centrifugal units. The resulting fragments were ligated using T4 ligase (0.1 units) at room temperature overnight. The DNA was precipitated from the mixture and re-suspended in sterile water (5 μL), which was used to transform *E. coli* strain DH5α cells by electroporation. The transformed cells were plated on LB/Kn (50 μg/mL) agar plates, and grown overnight at 37 °C. Four colonies were chosen at random and the plasmids purified using the Qiaprep Spin Miniprep Kit. Each of the four plasmids was digested with the *Nde*I and *Not*I restriction enzymes to screen for the presence of insert (i.e., the α-subunit) as indicated by the appropriate size band (250 bp) upon electrophoresis on a 1% agarose gel. Only one of the four plasmids contained the α-subunit. The plasmid was designated pET-24a(+)-CaaDα. Sequencing of the insert and the adjacent sequence verified that no mutations had been introduced.

The β-subunit was synthesized in two stages of PCR in a comparable manner using the oligonucleotides shown in Table S2. The most intense band (~250 bp in size) observed after the second stage was excised and purified using an Ultrafree-DA centrifugal unit. The PCR product and the pET-21a(+) vector were treated with the *Nde*I and *Not*I restriction enzymes and the products of these reactions purified and ligated as described above. The ligated DNA was precipitated and used to transform *E. coli* strain DH5α as described above. An aliquot of the transformed cells, after being treated as described above, was plated on LB/Ap (100 μg/mL) agar plates, and grown overnight at 37 °C. A total of 18 colonies were chosen at random and screened for the presence of a plasmid having the β-subunit insert, as indicated by the appropriate size band (~250 bp) upon electrophoresis on a 1% agarose gel. Four plasmids were identified that contained the insert. Sequencing of all four inserts (and

the adjacent sequences) uncovered two constructs (designated pET-21a-CaaDβ4 and pET-21a-CaaDβ6) without mutations.

Aliquots (0.5-1 μL) of each plasmid (pET-24a-CaaDα and pET-21a-CaaDβ4) were used to transform either *E. coli* strain BL21(DE3)pLysS or *E. coli* strain BL21-Gold(DE3)pLysS cells by electroporation following the manufacturer's directions. An aliquot (50-100 μL) was plated on LB/Kn (50 μg/mL)/Ap (100 μg/mL)/Cm (25 μg/mL) agar plates, and grown overnight at 37 °C. Only those colonies containing the two expression plasmids and the pLysS plasmid will grow on the triple antibiotic media (21).

*Overexpression and Purification of CaaD*. A 25 mL culture of LB containing both ampicillin (100 μg/mL) and kanamycin (50 μg/mL) was inoculated with a single colony of the expression strain and grown overnight at 37° C. Subsequently, a sufficient quantity of the overnight culture was used to inoculate 500 mL of LB/Kn/Ap medium in a 2 L Erlenmeyer flask to give an $OD_{600nm}$ reading of ~0.05. Cultures (6-2L flasks) were grown at 37° C to an $OD_{600nm}$ of ~0.6-0.8 with vigorous shaking and then induced with 0.5 mM IPTG. Additional aliquots of ampicillin (50 μg/mL) and kanamycin (25 μg/mL) were added to each culture upon induction in order to promote the retention of both expression plasmids (pET-21a-CaaDβ4 and pET-24a(+)-CaaDα). Cells were harvested after four hours of growth and stored at -80° C until use. Typically, 3 L of culture yields 12 g cells.

The purification of recombinant CaaD to near homogeneity was carried out by the following procedure. The buffers used in this purification are Buffer A: 50 mM Tris-$SO_4$, pH 8.0; Buffer B: 50 mM Tris-$SO_4$, 1 M $(NH_4)_2SO_4$, pH 8.0; and Buffer C: 20 mM $NaH_2PO_4$, pH 7.3. The cells (23.6 g) were thawed, resuspended in 3 volumes (71 mL) of buffer A, placed on ice, and initially disrupted by sonication at 50% output for one minute. Phenylmethylsulfonyl fluoride (0.5 mM) and 6-aminocaproic acid (1 mM) were added to the cell suspension. Subsequently, the cell suspension was subjected to further sonication for six 5-min intervals using 5-s pulses with a cycle time of 50%. After each 5-min interval, the cells were allowed to sit for 5

minutes. Protamine sulfate (0.47 g) was added to the lysate and the resulting mixture was stirred for 10 min at 4 °C, and then centrifuged at 19, 000 rpm for 30 min at 4 °C. The supernatant was removed, made 1 M in $(NH_4)_2SO_4$, and allowed to stir for 60 min at 4 °C.

After centrifugation at 17, 000 rpm for 30 min, the supernatant was loaded onto a Phenyl-5PW column, previously equilibrated in Buffer B. Protein was eluted using a 20-min wash of 100% Buffer B, followed by a 60 min linear $(NH_4)_2SO_4$ gradient (1.0-0.0 M) using Buffer A, and a final 10-min wash with 100% Buffer A. The flow rate was 5 mL/min. Fractions (7.5 mL) were collected and assayed for activity. Typically, the CaaD activity elutes 25 min after the initial wash. The most active fractions were pooled (~200 mL), concentrated, and exchanged into Buffer A using an Amicon stirred cell containing a YM-3 membrane. The resulting concentrate (~29 mL) was made 1 M in $(NH_4)_2SO_4$, and allowed to stir for 60 min at 4 °C. After centrifugation at 17, 000 rpm for 30 min, the supernatant was removed and passed through the Phenyl column using the conditions described above. The most pure and active fractions were pooled (~ 80 mL), concentrated to ~ 6 mL, and loaded onto a Sephadex G-200 column (2.5 × 100 cm), equilibrated with Buffer C. The flow rate was about 0.4 mL/min. Fractions were collected for 25 min each and contained about 10 mL. Typically, CaaD elutes after about 12 hours. The most active fractions were pooled and concentrated. The purity was judged to be >95% as assessed by SDS-PAGE. The overall yield of CaaD is typically 15-20 mg of homogenous protein per liter of culture. For this preparation, a 4-L culture gave 70 mg of protein.

The native molecular mass of CaaD was estimated by size exclusion chromatography on a Superose 12 column. The enzyme was chromatographed in 200 μL portions (1.3 mg/mL) on the gel filtration column equilibrated with 20 mM sodium phosphate buffer, pH 7.3, at a flow rate of 0.4 mL/min. The protein eluted at ~31 min.

*Construction of the β-P1A- and α–R11A-CaaD Mutants*. Two mutants of

CaaD were constructed by overlap extension PCR as described elsewhere (22). In the β-P1A mutant of CaaD, Pro-1 in the β-subunit was changed to an alanine, while the α-subunit remained unchanged. In the α–R11A mutant of CaaD, Arg-11 in the α-subunit was changed to an alanine, while the β-subunit was unchanged. The external primers for each mutant were oligonucleotides 5'-GCG GAT AAC AAT TCC CCT CT-3' (primer $A^m$) and 5'-CTC AGC TTC CTT TCG GGC TT-3' (primer $D^m$). Primer $A^m$ corresponds to a region 36 bases upstream of the *Nde*I restriction site in both the pET-21a(+) and pET-24a(+) plasmids. Primer $D^m$ corresponds to the complementary sequence of a region 20 bases downstream of the His-tag region in both the pET-21a(+) and pET-24a(+) plasmids. The internal primers for the β-P1A mutant were 5'-ATA CAT ATG GCC TTC ATC GAA-3' (primer $B^m$) and 5'-TTC GAT GAA GGC CAT ATG TAT-3' (primer $C^m$). The internal primers for the α-R11A mutant were 5'-CGT TAT GGG GCA ACA GAC GAA-3' (primer $B^m$) and 5'-TTC GTC TGT TGC CCC ATA ACG-3' (primer $C^m$). In each internal primer, the mutation is underlined and the remaining bases correspond to the coding sequence (primers $C^m$) or the complementary sequence (primer $B^m$) for the appropriate CaaD subunit gene.

For the construction of the $A^mB^m$ and $C^mD^m$ fragments for the β-P1A mutant, each PCR reaction contained 2.5 μL of 20 μM solutions of primers $A^m$ and $B^m$ or $C^m$ and $D^m$, 1 μL of a 10 mM stock of dNTPs, 5 μL of 10 × Taq reaction buffer, 3 μL of 25 mM $MgCl_2$, 5 μL of a 1:500 dilution of the pET-21a(+)-CaaDβ4 plasmid miniprep, 0.5 μL of Taq DNA polymerase, and a sufficient quantity of sterile water to make a final volume of 50 μL. The PCR protocol consisted of a 5-min incubation period at 94 °C, followed by 25 cycles of the PCR where each cycle contained a denaturation step at 94 °C for 60 s, an annealing step at 55 °C for 75 s, and an elongation step at 72 °C for 75 s, and a final 5-min incubation period at 72 °C. The amplified products were analyzed by electrophoresis on 1% agarose gels and purified using the Wizard PCR Preps DNA purification system. The $A^mD^m$ fragment was constructed similarly except that the $A^mB^m$ and $C^mD^m$ fragments (1 μL of each) were used as template. The

$A^mD^m$ fragment was electrophoresed, excised, and purified using the Ultrafree-DA centrifugal filter units. Digestion and purification of the $A^mD^m$ fragment and the pET-21a(+) vector were carried out as described above. Ligation was performed using the Quick Ligation Kit (New England Biolabs, Inc., Beverly, MA). The purified DNA was used to transform *E. coli* strain BL21-Gold(DE3) pLysS by electroporation. Transformed cells were plated onto LB/Ap (50 μg/mL)/Cm (25 μg/mL) and incubated overnight at 37 °C. Four colonies were selected at random and their plasmids were isolated for sequence analysis. Only one colony possessed the correct sequence for the desired mutant and was designated pET-21a(+)-CaaDβ-P1A.

For construction of the α-R11A mutant, Herculase DNA polymerase (Stratagene) replaced Taq DNA polymerase in the PCR reactions and $MgCl_2$ was not added. The PCR protocol consisted of a 2-min incubation period at 92 °C, followed by 30 cycles of the PCR where each cycle had a 30-sec denaturation step at 92 °C, a 30-sec annealing step at 60 °C, and a 60-sec elongation step at 72 °C, and a final 10-min incubation period at 72 °C. Purification of the $A^mB^m$ and $C^mD^m$ fragments and construction of the $A^mD^m$ fragment was carried out as described above. Subsequently, the pET-24a(+) vector and the $A^mD^m$ fragment were treated with the appropriate restriction enzymes at 37 °C for 6 hr and purified using gel electrophoresis and the Ultrafree-DA centrifugal filter units. Ligation and transformations were carried out as described for the construction of the β-P1A mutant. The purified DNA was used to transform *E. coli* DH5α by electroporation. Transformed cells were plated onto LB/Kn (50 μg/mL) and incubated overnight at 37 °C. Two colonies were selected at random for sequencing. One colony, designated pET-24a(+)-CaaDα-R11A2, had only the desired mutation.

*Overexpression and Purification of the β-P1A- and the α-R11A CaaD Mutants*. For expression, each mutant plasmid was transformed into *E. coli* strain BL21-Gold(DE3) pLysS by electroporation along with the appropriate wild-type partner plasmid. Thus, for expression of the β-P1A mutant, both the pET-21a(+)-CaaDβ-P1A and pET-24a(+)-CaaDα vectors were used to transform cells. For

expression of the α-R11A mutant, both the pET-21a(+)-CaaDβ4 and pET-24a(+)-CaaDα-R11A2 vectors were used to transform cells. Cells were selected for the presence of the two plasmids by growth on LB plates containing the triple antibiotics as described above. Cells expressing the mutant proteins were grown using the protocol described for the wild type enzyme. Subsequently, the mutant proteins were purified following the procedure used for the wild type CaaD. Typically, a 3L culture of the β-P1A mutant yields 10 g of cells and 50 mg of ~95% pure protein as assessed by SDS-PAGE. A 3L culture of the α-R11A mutant yields 10 g of cells and 40 mg of ~90% pure protein as assessed by SDS-PAGE.

The native molecular masses of the two mutants were estimated by size exclusion chromatography on a Superose 12 column. Both enzymes were chromatographed in 200 μL portions (0.7 mg/mL) on the gel filtration column equilibrated with 20 mM sodium phosphate buffer, pH 7.3, at a flow rate of 0.4 mL/min. The β-P1A CaaD mutant eluted at ~32 min, while the α-R11A CaaD mutant eluted at ~31 min.

*Mass Spectrometric Characterization of CaaD and the Mutants*. The subunit masses of CaaD and the two mutants were determined using an LCQ electrospray ion trap mass spectrometer (ThermoFinnigan, San Jose, CA), housed in the Analytical Instrumentation Facility Core in the College of Pharmacy at the University of Texas at Austin. The protein samples (~1 mg/ml in 100 mM $(NH_4)_2CO_3$ buffer, pH 8.0) were mixed in a 1:10 ratio with a solution of 50% (v/v) methanol in water containing 5% acetic acid. The samples were infused into the mass spectrometer at 20 μl/min and spectra were acquired continuously in the centroid mode over the mass range of *m/z* 350-2000. The deconvoluted masses were calculated using ThermoFinnigan's Xcalibur 1.3 BioWorks 3.0 software with approximately 40 spectra averaged.

The observed monomer masses for the α- and β-subunits of CaaD were 8342 (calc. 8344) and 7505 (calc. 7507), respectively. The observed monomer masses for the α- and β-subunits of β–P1A-CaaD were 8342 (calc. 8344) and 7479 (calc. 7481),

respectively. The observed monomer masses for the α- and β-subunits of α–R11A-CaaD were 8257 (calc. 8258) and 7506 (calc. 7507), respectively.

*Enzymatic Assay and Kinetic Studies of CaaD*. The kinetic assays were performed in 20 mM $Na_2HPO_4$ buffer, pH 9.0, observing the decrease in absorbance at 224 nm corresponding to the hydration of *trans*-3-chloroacrylate (**1**, $\varepsilon = 4900\ M^{-1}\ cm^{-1}$) and *trans*-3-bromoacrylate (**11**, $\varepsilon = 9700\ M^{-1}\ cm^{-1}$). An aliquot of CaaD (100 μL of a 13.2 mg/mL solution) was diluted into buffer (40 mL) and incubated for 1 h. Subsequently, a 1 mL-portion of the enzyme was transferred to a cuvette and assayed by the addition of a small quantity of substrate (**1** or **11**) from a 5 mM or 50 mM stock solution (in 100 mM $Na_2HPO_4$ buffer, pH 9.2). The addition of **1** or **11** (as their free acids) to the 100 mM phosphate buffer adjusted the pH of the stock solution to about 7.0. The 5 mM stock solution was made by dilution of an aliquot of the 50 mM stock solution into 100 mM $NaH_2PO_4$ buffer, pH 7.3. The concentrations of substrate used in the assay ranged from 10-150 μM.

The hydration of 2-oxo-3-pentynoate (**7**) by CaaD was monitored by following the formation of acetopyruvate (**8**) at 294 nm ($\varepsilon = 7{,}000\ M^{-1}\ cm^{-1}$) in 20 mM sodium phosphate buffer (pH 9.0) (23). An aliquot of CaaD (500 μL of a 2.9 mg/mL solution) was diluted into buffer (50 mL) and incubated for 1 h. The assay was initiated by the addition of a small quantity of **7** from either a 5.95 mM or 59.5 mM stock solution. The stock solution (59.5 mM) was made by dissolving the appropriate amount of 2-oxo-3-pentynoate in 100 mM $Na_2HPO_4$ buffer, pH 9.2. The addition of **7** (as the free acid) adjusted the pH of the buffer to about 7. The 5.95 mM stock solution was made by dilution of an aliquot of the 59.5 mM into 100 mM $NaH_2PO_4$ buffer, pH 7.3. The concentrations of **7** used in the assay ranged from 6-595 μM.

*$^1H$ NMR Spectroscopic Detection of **3***. An NMR tube contained 100 mM $Na_2HPO_4$ buffer (0.54 mL, pH ~9.2) and a solution of **1** (4 mg, 0.04 mmol) dissolved in DMSO-$d_6$ (30 μL). The final pH of the solution was 6.8. Subsequently, an aliquot of CaaD (60 μL of a 6.7 mg/mL solution in 20 mM sodium phosphate buffer, pH 7.3)

was added to the reaction mixture. The concentration of **1** in the NMR tube was 62 mM. The first $^{1}$NMR spectrum was recorded 5 min after the addition of the enzyme and every 5 min thereafter until the reaction was completed (30 min). **3**: $^{1}$H NMR ($H_2O$, 500 MHz) δ 3.20 (2H, d, $J$ = 3 Hz, H-2), 9.50 (1H, t, H-3). The spectra also showed signals corresponding to acetaldehyde (resulting from the non-enzymatic decarboxylation of **3**), and to the hydrates of **3** [$^{1}$H NMR ($H_2O$, 500 MHz) δ 2.34 (2H, d, $J$ = 5.4 Hz, H-2), 5.15 (1H, t, H-3)] and acetaldehyde (24).

*$^{13}$C NMR Spectroscopic Detection of **3**.* An NMR tube contained 200 mM $Na_2HPO_4$ buffer (0.3 mL, pH ~8.5), 1M NaOH (175 μL), $H_2O$ (10 μL), and a solution of **1** (20 mg, 0.19 mmol) dissolved in DMSO-$d_6$ (30 μL). The final pH of the solution was 6.8. After sitting at 4 °C overnight, an aliquot of CaaD (115 μL of a 6.7 mg/mL solution in 20 mM sodium phosphate buffer, pH 7.3) was added to the reaction mixture. The concentration of **1** in the NMR tube was 310 mM. The initial $^{13}$C NMR spectrum was obtained 25 min after the addition of CaaD. **3**: $^{13}$C NMR ($H_2O$, 500 MHz) δ 46.2 (C-2), 175.4 (C-1), 204.2 (C-3). The spectra also showed evidence for the presence of acetaldehyde (resulting from the non-enzymatic decarboxylation of **3**), and the hydrates of **3** [$^{13}$C NMR ($H_2O$, 500 MHz) δ 53.6 (C-2), 90.3 (C-3), 178.5 (C-1)] and acetaldehyde (24).

*UV and $^{1}$H NMR Spectroscopic Detection of Acetopyruvate (**8**) in the CaaD-catalyzed Hydration of 2-Oxo-3-pentynoate (**7**).* In the UV spectrophotometric experiment, the conversion of **7** to **8** was performed by adding a quantity of CaaD (5 μL of a 7.9 mg/mL solution) to a cuvette containing buffer (1 mL of 20 mM $Na_2HPO_4$ buffer, pH 9) and an aliquot of **7** (2 μL from a 59.5 mM solution in 100 mM $Na_2HPO_4$ buffer, pH 9.2). Spectra were recorded every 30 s until the reaction neared completion (approx. 15 min). The region from 200-350 nm was examined using the Agilent 8453 Diode Array spectrophotometer at 27 °C. For the $^{1}$H NMR experiment, **7** (4 mg, 36 mmol) dissolved in DMSO-$d_6$ (30 μL), was added to 100 mM $Na_2HPO_4$ buffer, pH 9.2 (0.6 mL) in an NMR tube. The addition of **7** adjusted the pH of the buffer to 6.5. Subsequently, an aliquot of CaaD (200 μL of a 11 mg/mL solution) was

added to the mixture to initiate the reaction. NMR spectra were recorded at 3.5 min and 10 min after the addition of enzyme, when the reaction was completed. The $^1$H NMR spectra show the presence of signals corresponding to **8**, its hydrate **12** (25), and the enol isomer, **13** (*26*). $^1$H NMR (100 mM $Na_2HPO_4$, 500 MHz) δ 2.02 (s, 3H, $CH_3$ of **12**), 2.05 (s, 3H, $CH_3$ of **13**), 2.12 (s, 3H, $CH_3$ of **8**), 2.88 (s, 2H, $CH_2$ of **12**), 3.79 (s, 2H, $CH_2$ of **8**), 5.94 (s, 1H, enol proton of **13**).

*Detection of the β-P1A- and α–R11A-CaaD Mutant Activity*. Activities for the β-P1A- and α–R11A-CaaD mutants could not be determined by UV spectroscopy because the large amount of protein required interfered with the UV absorbance of substrate. Hence, activity was detected using $^1$H NMR spectroscopy as follows. To an NMR tube containing 100 mM $Na_2HPO_4$ buffer (0.6 mL, pH ~9.2) was added a solution of 3-chloroacrylic acid (**1**, 4 mg, 0.04 mmol) dissolved in DMSO-$d_6$ (30 μL). The addition of **1** to the buffer lowered the pH to ~6.8. Subsequently, an aliquot of β-P1A-CaaD or α–R11A-CaaD (100 μL of a 7.4 mg/mL solution) was added to separate NMR tubes. The reaction mixture was allowed to incubate for 24 hrs. The final pH was 6.65 (β-P1A-CaaD) or 6.53 (α–R11A-CaaD). A $^1$H NMR spectrum revealed the presence of acetaldehyde (resulting from the non-enzymatic decarboxylation of **3**) and its hydrate in the NMR tube containing α–R11A-CaaD, but no products in the NMR tube containing β-P1A-CaaD.

*Kinetics of Irreversible Inhibition.* A quantity (100 μL) of CaaD (5.7 mg/mL) was diluted 10-fold into 20 mM sodium phosphate buffer (pH 7.3) to give a final concentration of 36 μM (based on dimer molecular mass). The diluted enzyme was divided into 150 μL quantities, placed in 1.5 mL eppendorf micro test tubes, and equilibrated at 23 °C for 1 hr before the addition of inhibitor. A 20 μL-aliquot was removed just prior to the addition of inhibitor and assayed for activity. This activity represents 100% activity at time zero. Subsequently, varying amounts of **9** or **10** (0–40 μM) and a sufficient quantity of buffer (100 mM $Na_2HPO_4$ buffer, pH 7.3) were added to the incubation mixtures to make a final total volume of 140.4 μL. Aliquots (20 μL) were removed from these mixtures immediately after the addition of inhibitor

(~ 5 s) and at 5-s intervals thereafter, diluted into 20 mM sodium phosphate buffer (1 mL, pH 9.0), and assayed for residual activity. The assay was initiated by the addition of **11** (150 μM). Each time point (shown in Figure 31) represents the average of three different runs. Stock solutions (50 mM) of **9** and **10** were made up in 100 mM $Na_2HPO_4$ buffer (pH 9.2). The addition of the inhibitor resulted in a final pH of ~7. The stock solutions were diluted with a sufficient volume of 100 mM sodium phosphate buffer (pH 7.3) to make 5 mM and 0.5 mM solutions of inhibitor.

*Protection of CaaD from Inhibition by **9** and **10**.* The protection of CaaD from inactivation by **9** and **10** was carried out as follows. Quantities of CaaD (150 μL) were placed in 1.5 mL eppendorf micro test tubes and the initial time point was obtained as described above. Aliquots (10.4 μL) of a solution containing either **9** (0.25 mM or 0.5 mM) and **11** (14 mM, 65 mM, and 138 mM) or **10** (0.25 or 0.5 mM) and **11** (14 mM, 65 mM, and 138 mM) were then added to the eppendorf tubes. This resulted in final concentrations of either **9** (20 μM or 40 μM) or **10** (20 μM or 40 μM) and **11** (1 mM, 5 mM, and 10 mM). Aliquots (20 μL) were removed from these mixtures as described above, diluted into 20 mM sodium phosphate buffer (1 mL, pH 9.0, a 50-fold dilution), and assayed for residual activity using the quantity of **11** present in the incubation mixture.

*Bromide Elimination from **9** in the Presence of CaaD.* The release of bromide ion was monitored using an Accumet AP63 pH/millivolt/ion meter equipped with an Accumet bromide combination ion selective electrode (Fisher Scientific Inc.). The amount of bromide in solution was determined from a standard curve following the manufacturer's directions. The reaction mixtures (total volume of 3 mL) contained varying amounts of **9** (0 μM, 25 μM, 150 μM, or 300 μM) made up as described above and CaaD (297 μL from a 21.5 mg/mL solution resulting in a final concentration of 127 μM). The reactions were initiated by the addition of inhibitor. An aliquot (60 μL) of a 5M $NaNO_3$ solution was added to each reaction mixture to maintain a constant ionic strength.

In order to determine the response time of the electrode, the amount of bromide ion was monitored in a reaction mixture containing 20 mM $NaH_2PO_4$ buffer, pH 9.0, (1.5 mL), $NaNO_3$ (30 μL of a 5 M solution), and NaBr (30 μL of a 5 mM solution) under the conditions described above. The final concentration of bromide ion in solution (50 μM) was determined from a standard curve to be present in solution after ~3 min.

*Irreversibility of the Inactivation*. The irreversibility of the reaction was established using both **9** and **10**. CaaD (1.53 μM based on the molecular weight of the dimer) was incubated with an excess of **9** or **10** (50 μM) in 5 mL of 20 mM $Na_2HPO_4$ buffer (pH ~9) for 20 min at 23 °C. The final pH of the solution was 8.0. In a separate control reaction, an identical quantity of enzyme was incubated without inhibitor under identical conditions. The treated samples had no activity after 20 min. The three samples were dialyzed against 20 mM $Na_2HPO_4$ buffer, pH 9.0 at 22 °C. After 63 h, the control sample lost ~40% of its original activity, while the treated samples did not regain any activity.

*Trypsin Digestion of CaaD and CaaD-modified by **9***. Each sample contained ~1 mg of CaaD (125 μl of a 7.9 mg/mL solution) and a sufficient quantity of 20 mM sodium phosphate buffer, pH 7.3, to give a final volume of ~0.5 mL. One sample was treated with **9** (13 μL from a 44.8 mM stock solution of **9** in 100 mM $Na_2HPO_4$ buffer). Both samples were incubated at 4 °C for ~12 hrs and passed through separate PD-10 Sephadex columns equilibrated in 100 mM $(NH_4)_2CO_3$ buffer, pH 8. Fractions (~0.5 mL) were collected, and protein was identified by the Waddell method (17). Subsequently, both samples were treated with sequencing grade trypsin by a modification of a literature procedure (26). Accordingly, aliquots (300 μL) of CaaD and modified CaaD were made 1M in guanidine hydrochloride, 5 mM in $CaCl_2$, and 1 mM in dithiothreitol (DTT). The sequencing grade trypsin (100 μg) was reconstituted according to the manufacturer's instructions in 1 mM HCl (10 μL) and a 2-μL aliquot was added to each protein sample. The samples were incubated at 37 °C overnight.

Subsequently, the mixtures were subjected to matrix assisted laser desorption-ionization (MALDI) mass spectral analysis as described below.

*Mass Spectrometry of the Digested Proteins*. The trypsin-digested CaaD samples were analyzed on the delayed extraction Voyager-DE PRO MALDI-TOF instrument (PerSeptive Biosystems, Framingham, MA) operating in the positive ion mode. The instrument is equipped with a 337 nm nitrogen laser with a 20 Hz firing rate. The matrix consisted of α-cyano-4-hydroxycinnamic acid from the Sequazyme Peptide Mass Standards Kit (PerSeptive Biosystems) dissolved in 50% (v/v) acetonitrile in water with 0.1 % trifluoroacetic acid (TFA). Samples (1 mg/mL in 100 mM $(NH_4)_2CO_3$ buffer, pH 8) were mixed with the matrix in a 1:10 ratio. A 1 μl aliquot was removed and drop dried on a stainless steel target. Close external calibration was performed using Calibration Mixture 1 (Sequazyme Peptide Mass Standards Kit) made up according to the manufacturer's directions. The 2.0 m reflector detector was used with the low mass gate set at *m/z* 500, and spectra were acquired over the range of *m/z* 500-3500. The instrument parameter file "angiotensin_reflector" was used for data acquisition, with 20 kV accelerating voltage, 74% grid voltage, 0.002% guide wire voltage and 75 nsec delay time. Up to 400 shots were averaged for the spectrum with typical resolution of 8000. MALDI-Post Source Decay (PSD) spectra were composite spectra generated from a set of 11 mirror ratios acquired with the instrument parameter file "angiotensin_PSD". Up to 400 shots were averaged for each segment. Air was used for collision induced dissociation at masses less than *m/z* 220.

## RESULTS

*Construction, Expression, and Characterization of CaaD.* CaaD has been constitutively expressed in both *E. coli* and *Pseudomonas* sp. strain GJ1 (3). In order to optimize expression of the unlabeled enzyme as well as isotopically labeled enzyme for NMR studies, and to facilitate mutagenesis, the genes for the α– and β– subunits of CaaD from *Pseudomonas pavonaceae* strain 170 were synthesized in separate reactions using a PCR-based gene synthesis strategy (18-20), and cloned into the T7 expression system. Accordingly, the gene for the α-subunit was divided into five long oligonucleotides with short overlapping regions, while the gene for the β-subunit was divided into four long oligonucleotides with short overlapping regions. These overlaps serve as primers for the extension of the long oligonucleotides, resulting in their fusion in the course of the PCR. The full-length genes are then amplified by the inclusion of short primers (corresponding to the termini). Subsequently, the two gene fragments were cloned into separate plasmids, which were, in turn, transformed into the expression strain. The sequences of the subunits were confirmed by DNA sequencing. CaaD was purified to near homogeneity by hydrophobic interaction chromatography followed by gel filtration. Typically, the expression system yielded 17-20 mg of homogeneous protein (~95% pure as estimated by SDS-PAGE) per liter of culture.

The purified CaaD was analyzed by electrospray ionization mass spectrometry (ESI-MS) and gel filtration chromatography. A sample generates two major peaks in the mass spectrum (after deconvolution) that correspond to the expected masses of the α– and β– subunits. The observed masses of the two peaks indicate that the amino-terminal proline is not blocked in either subunit by the initiating methionine. The native molecular mass (as estimated on a gel filtration column) is comparable to the previously reported mass, suggesting that CaaD is a heterohexamer in solution (3).

*Construction, Expression, and Characterization of CaaD Mutants*. The β-P1A and the α-R11A mutants of CaaD were constructed by overlap extension PCR

using the appropriate synthetic gene (in either the pET-21a or pET-24a plasmid) as template (22). The resulting plasmid containing the mutant gene was co-transformed into the *E. coli* strain BL21-Gold(DE3)pLysS along with the appropriate wild type plasmid. The sequence of each mutant was confirmed by DNA sequencing. Subsequently, the mutant proteins were expressed and purified following the protocol developed for the wild type enzyme. The expression system yielded 10-15 mg for the β-P1A mutant and 15-20 mg for the α-R11A mutant (per liter of culture). Analysis of the mutant proteins by ESI-MS and gel filtration chromatography indicated that the α– and β–subunits of the mutants had the expected masses (and were not blocked by the initiating methionine) and that the mutants were heterohexamers in solution.

*Kinetic Properties of CaaD, β-P1A-CaaD, and the α-R11A-CaaD*. The activity of CaaD was previously determined by a colorimetric assay that monitored the release of the halide (3). It was necessary to have a more direct assay for CaaD activity in order to facilitate inhibition studies, pH rate studies, and kinetic studies of CaaD mutants. Hence, the activity of CaaD was monitored by following the decrease in the absorbance of the substrate at 224 nm. This assay was used to measure kinetic parameters for CaaD using **1** and **11** (Table 2). The two substrates have comparable $K_m$ values, while the $k_{cat}$ value for **11** is 1.3-fold greater than that for **1**. The increase in $k_{cat}$ may reflect the fact that bromide is a better leaving group than chloride (if the loss of halide reflects the rate-determining step). The overall effect of the increase in $k_{cat}$ is a slightly higher (~1.2-fold) $k_{cat}/K_m$ value for **11** (compared to **1**). A $K_m$ of 0.19 mM and a $k_{cat}$ of 6.4 $s^{-1}$ were previously reported for CaaD using **1** (3). The differences (most notably in $K_m$) are likely a function of the two different assays.

Both the β-P1A and the α-R11A mutants of CaaD were reported to be inactive using **1** (3). The β-P1A mutant showed a small amount of activity using **11**, but the α-R11A mutant did not (3). In the spectrophotometric assay, activity was not detected for either mutant **1** or **11**. However, a 24 hr incubation of both mutant proteins (in separate reactions) with **1** showed that the α-R11A mutant retained a

small amount of activity (as determined by the presence of the characteristic signals in a $^{1}$H NMR spectrum) while the β-P1A mutant had no detectable activity.

*$^{1}$H and $^{13}$C NMR Characterization of the CaaD-catalyzed Reaction.* In order to directly observe the product, the CaaD-catalyzed reaction was monitored by $^{1}$H and $^{13}$C NMR spectroscopy (Figure 29). Incubation of CaaD and **1** for 5 min resulted in the spectrum shown in Figure 29A. The doublets centered at 6.09 and 6.92 ppm correspond to the protons at C-2 and C-3 of **1**, respectively. Malonate semialdehyde (**3**) is responsible for the doublet at 3.20 ppm and the triplet at 9.50 ppm, which correspond to the protons at C-2 and C-3, respectively. In addition, signals corresponding to the hydrate of **3** appear as a doublet at 2.34 ppm and a triplet at 5.15 ppm, corresponding to the protons at C-2 and C-3, respectively. Finally, the spectrum shows much smaller signals readily assigned to acetaldehyde and its hydrate (25). After 20 min (Figure 29B), the signals corresponding to **1** have largely collapsed while those corresponding to **3** and its hydrate remain. Over time, the signals corresponding to acetaldehyde and its hydrate grow in intensity. In view of the significantly slower rate of formation of acetaldehyde, it can be reasonably concluded that it results from the non-enzymatic decarboxylation of **3**. The $^{13}$C NMR spectrum (recorded 25 min after the addition of CaaD to **1**) shows signals that can be most reasonably assigned to **3** and its hydrate.

*Hydration of 2-Oxo-3-pentynoate (7) by CaaD*. 2-Oxo-3-pentynoate (**7**) was previously synthesized and found to be a potent active site-directed irreversible inhibitor of 4-OT (12). In view of the sequence similarities between CaaD and 4-OT, and the importance of the β–Pro-1 to the enzyme-catalyzed reaction (3), **7** was initially examined as a potential affinity label of CaaD. Incubation of **7** with CaaD resulted in a decrease in the absorbance corresponding to **7** ($\lambda_{max}$ = 234 nm), accompanied by the appearance of a new absorbance peak at 294 nm, which corresponds to acetopyruvate (**8**) (Figure 30). In addition to the characteristic $\lambda_{max}$, the identity of **8** was further confirmed by $^{1}$H NMR spectroscopy. The $^{1}$H NMR spectrum showed signals consistent with the structure of **8** (26), as well as two

additional species, the hydrate **12** (26), and the enol **13** (26) (Scheme 15). No significant inactivation of CaaD by **7** was observed in these experiments. In the absence of CaaD, **7** is stable for several hours in solution and does not decompose to **8** (12).

The kinetic parameters for the conversion of **7** to **8** were also determined and are summarized in Table 2. The values of $k_{cat}$ are 5 to 7-fold lower than those determined for **1** and **11**, respectively, while the $K_m$ is about 3-fold higher. The net effect is a modest decrease (19-22-fold) in the $k_{cat}/K_m$ (when compared with those determined for **1** and **11**). Presumably the additional carbonyl group as well as the sp hybridization of **7** is detrimental to its binding (resulting in a higher $K_m$), which could misalign the substrate with respect to the catalytic groups of CaaD and cause the observed decrease in $k_{cat}$.

*Inactivation of CaaD by **9** or **10**.* Incubation of CaaD with either **9** or **10** results in a rapid irreversible inactivation of the enzyme (Figure 31). The enzyme loses ~80 % of its activity in ~10 s when incubated with a stoichiometric excess (10% based on the dimer molecular mass of CaaD) of **9** (Figure 31A). Under the same conditions, CaaD loses 80% of its activity in ~5 s when incubated with **10** (Figure 31B). The rapid loss of activity did not permit the collection of sufficiently precise data to assess whether the loss of activity occurs in a time-dependent manner or whether the observed differences in the inactivation times are meaningful. Various attempts to slow the inactivation process by lowering the temperature to 4 °C or by varying the pH over the range of 5.5-9.0 were not successful. Under all conditions, CaaD was rapidly inactivated. Dialysis (63 h) of CaaD inactivated by either **9** or **10** does not result in any reactivation of CaaD, consistent with the formation of a covalent bond between the enzyme and a species resulting from **9** (or **10**).

Binding at the active site of CaaD was suggested by the observation that **11**, if added with either inhibitor, protects CaaD from inactivation. The results are summarized in Table 3 and show a clear trend. A lower concentration of **11** (1 mM which is ~25-fold higher than the $K_m$ of **11**) prolongs the enzyme's lifetime by a few

seconds while higher concentrations (5 and 10 mM which are 200-250 higher than the $K_m$ of **11**) prolong it by tens of seconds. Presumably, when the enzyme converts all of substrate **11** present in the incubation mixtures to **3**, the protection is lost.

*Bromide Elimination from **9** in the Presence of CaaD*. The elimination of bromide ions was observed for increasing concentrations (25 μM, 150 μM, and 300 μM) of **9**, when incubated with a fixed concentration of CaaD (127 μM) (Figure 32) at 24 °C and pH 8.0. In all cases, the majority of bromide ion release is complete within 3 min and is dependent on the presence of CaaD.

In the presence of a stoichiometric amount of **9**, CaaD is completely inactivated under these conditions within 10 s. This observation initially suggested that the observed bromide ion elimination and the inactivation of CaaD by **9** might not be a coupled process. However, a control experiment indicated that the rate of bromide ion release cannot be accurately determined under these conditions. In this experiment, the concentration of bromide ion in solution was monitored using a 50 μM solution of NaBr. After ~3 min, a total of 50 μM of bromide ion was achieved in solution (as detected by the electrode). Thus, the rate of CaaD inactivation by **9** is too fast to obtain accurate data that can be correlated with the rate of bromide ion elimination. Nonetheless, there is a correlation between inactivation and the release of bromide ion. At sub-stoichiometric concentrations of **9** (compared with the concentration of CaaD), bromide ion release corresponds to the concentration of **9**. When **9** is in a stoichiometric excess, bromide ion release corresponds to the concentration of CaaD.

*Identification of the Modified Peptide by Mass Spectrometry*. ESI mass spectral analysis of the intact α- and β–subunits of CaaD, after its incubation with **9**, indicated that only the β–subunit was modified by the inhibitor. The molecular mass of the β–subunit increases from 7505 ± 3 Da to 7590 ± 3 Da after incubation with **9**, while the mass of the α-subunit remains unchanged at 8342 ± 3 Da. The increase in mass is due to the covalent attachment of a species with a mass of 85 ± 3 Da.

In order to identify the site of attachment, CaaD was inactivated by **9**, (designated the treated sample), incubated overnight with trypsin, and analyzed by MALDI-TOF mass spectrometry. A control sample was made up similarly, but excluding **9** from the mixture. The site of covalent attachment was determined by comparative MALDI-TOF of these tryptic digests followed by selected fragmentation of the peptides of interest (27). The results (for the β-subunit) from the mass spectral analysis of each sample are summarized in Table 3 and show one major new peak appearing in the spectrum of the treated sample.

In the MALDI spectrum of the control sample (Figure 33A), there is a prominent peak at *m/z* 1200.63 ± 0.06, which corresponds to the calculated mass of the N-terminal sequence (PFIECHIATGL) of the β–subunit of CaaD[2]. This peak is not present in the MALDI spectrum of the treated sample (Figure 33B), but there is a new pair of peaks at *m/z* 1242.68 ± 0.06 and 1244.75 ± 0.12. These masses correspond to the N-terminal sequence (PFIECHIATGL) of the β–subunit of CaaD modified by adducts of 42 and 44 Da, respectively[3].

The ions corresponding to the selected peaks in the spectra of the control and the treated samples were subjected to MALDI-PSD analysis in order to determine the location of the modified residue (27). The MALDI-PSD spectrum of the peptide at *m/z* 1200 in the control sample, show a series of b ions, which have a charge on the N-terminal side of the fragmented peptide (Figure 34A). The MALDI-PSD spectra of the peptides in the treated sample show a similar b ion series, but now each ion is shifted by a mass of +42 (Figure 34B)[3]. Since all the $b_1$ ions are shifted by +42, the modification is on the N-terminal proline. While the $b_1$ ion, corresponding to the fragmentation of Pro-1, is not normally seen in PSD spectra, a $b_1$ +42 ion is seen at *m/z* 140.2 in the spectra of the treated sample. In the spectrum of the intact protein, there was a mass shift of ~85 Da, while in the digest, a mass shift of +42/+44 is seen. The discrepancy is probably due to decarboxylation of the adduct during the MALDI deposition or ionization process.

## DISCUSSION

Many synthetic halogenated hydrocarbons are environmental pollutants that frequently cause adverse health effects in humans and animals (28). Remarkably, microorganisms have evolved to use these compounds as sources of carbon and energy, converting them to carbon dioxide and water, thereby rendering such toxic substances less harmful. In most cases, this capability stems from the presence of specialized enzymes known as dehalogenases, which catalyze the cleavage of the carbon-halogen bond. There is much interest in the mechanisms of these enzymes because of their ability to catalyze a chemically difficult reaction, their potential use in bioremediation efforts, and their possible recent origin (28).

The majority of known microbial hydrolytic dehalogenases can be categorized into two broad groups, based on the whether the enzyme processes a halogenated aliphatic compound (29, 30) or a halogenated aromatic compound (31). Representative dehalogenases from both groups have been extensively studied, several crystal structures have been obtained, and a considerable body of mechanistic evidence has accumulated (28-31). These enzymes proceed through a two-step process. In the first step, an aspartate group attacks the substrate at the carbon to which the halogen is bound, resulting in loss of the halide and the formation of an alkyl (or aryl) enzyme intermediate. Subsequently, this intermediate is hydrolyzed by water. The catalytic aspartate is part of a catalytic triad or dyad, while the departure of the halide is sometimes facilitated by the presence of a halide-binding pocket (28-31).

While haloaromatic and haloalkyl dehalogenases are well characterized, haloalkene dehalogenases are rare and the catalytic strategies used by these enzymes are poorly understood. The only known haloalkene dehalogenases are CaaD and a *cis*-3-chloroacrylic acid dehalogenase (28). The recent characterization of CaaD, including sequence analysis and site-directed mutagenesis studies, demonstrated clear

mechanistic and structural differences between it and all previously characterized dehalogenases, and placed CaaD in the 4-OT family of enzymes (3).

The 4-OT family of enzymes is one of the three known families that comprise the tautomerase superfamily (32, 33). The other two families are represented by 5-(carboxymethyl)-2-hydroxymuconate isomerase (CHMI), another bacterial isomerase, and macrophage migration inhibitory factor (MIF), a mammalian cytokine with a phenylpyruvate tautomerase activity. The members of this superfamily are homologous proteins that share two major traits - they are constructed from a common β–α–β structural unit and they have an amino-terminal proline. In the title enzyme of each family, Pro-1 serves as a general base in a keto-enol tautomerization reaction. In the 4-OT family, the monomer of each known member encodes the characteristic β–α–β unit, and these monomers form stable hexamers (4-OT and YwhB) or dimers (YdcE, a 4-OT homologue found in *E. coli*) (33). While the α- and β-subunits of CaaD are grouped within one of the five subfamilies in the 4-OT family, the characteristics of the subfamily are not yet well defined (33).

In this context, the discovery of CaaD is highly significant because it represents a new activity (hydratase) as well as a new quaternary structure (heterohexamer) in the 4-OT family. The product of the reaction, **3**, was previously identified (and quantified) by its phenylhydrazone (1) and formazan derivatives (2). Derivatization was necessary due to the facile decarboxylation of **3**. In order to circumvent this problem and to directly detect **3** (and other possible products), the reaction was monitored by $^{1}$H NMR spectroscopy. In addition to the hydrate of **3** (which is readily stabilized by a 6-member ring), the NMR spectrum showed a small amount of acetaldehyde, resulting from the facile decarboxylation of **3** (1, 2). The slow rate of formation of the acetaldehyde is consistent with the non-enzymatic decay of **3**. No other products were present. On the basis of these observations, it is concluded that CaaD catalyzes primarily a hydration reaction to produce **3**, and it does not catalyze the decarboxylation of **3**.

Previous work also identified Pro-1 and Arg-11 as critical residues for the dehalogenation reaction and resulted in two proposals for the catalytic mechanism (3). One mechanism (Scheme 14) involves the activation of water by Pro-1 (acting as a general base) for attack at C-3 and the subsequent protonation at C-2 by Pro-1 (now functioning as a general acid). An alternate mechanism implicates an unidentified residue in the activation of a water molecule while the proton at C-2 is derived from Pro-1, functioning only as a general acid. In both mechanisms, Arg-11 plays a role in binding and perhaps catalysis: it interacts with the carboxylate group of the substrate and may function as an electron sink to facilitate catalysis. Electron density is drawn away from C-3 by the interaction with Arg-11, which generates a partial positive charge.

Whether Pro-1 functions as a general base or acid catalyst, our mutagenesis results further reinforce the critical nature of Pro-1 in the CaaD-catalyzed reaction. It was not possible to quantify the enzymatic activity of the β–P1A mutant (as well as the α–R11A mutant) using a more sensitive UV assay, which measures the depletion of substrate at 224 nm. However, it was determined by NMR spectroscopy that incubation of **1** and the β–P1A mutant for 24 hr did not produce acetaldehyde, which would result from the non-enzymatic decay of **3**. In 24 h, the α–R11A mutant generated a small quantity of acetaldehyde, suggesting, in turn, that a small amount of **3** had formed. In the absence of a crystal structure, it is not clear whether the inactivity of the β–P1A mutant and the diminished activity of the α–R11A are due to structural perturbations in addition to the absence of essential catalytic groups. Interestingly, the R11A mutation of 4-OT is significantly more detrimental to catalytic activity than is the P1A mutation, primarily due to an 8.6-fold increase in $K_m$ (8, 34). The mutations did not result in major structural perturbations to the 4-OT structure (8, 34).

It is noteworthy that the reaction catalyzed by CaaD is a particularly difficult one, as has been reported for other enzyme-catalyzed dehalogenation reactions. In 0.5 M aqueous NaOH at 60 °C, about 10% of the chloride is removed from **1** in 24 hr

(35). The kinetic analysis indicates that CaaD accelerates this process considerably and that **11** is a slightly better substrate for CaaD than is **1**. This could reflect the fact that C-3 is more susceptible to nucleophilic attack by water due to the greater activation at this position by the bromide. It is not likely due to a difference in the leaving group ability of the bromide ion versus the chloride ion (bromide:chloride 1.0:0.02) (36), as the 50-fold difference suggests that there would be a more significant difference in the rates.

CaaD will also process three acetylene compounds to products, which have different consequences for the enzyme. The first compound, 2-oxo-3-pentynoate (**7**), an irreversible inhibitor of 4-OT, functions as a substrate for CaaD, and is converted to acetopyruvate, which does not inactivate the enzyme. The most reasonable scenario for the formation of **8** from **7** is shown in Scheme 16. The CaaD-catalyzed addition of water to **7** results in the formation of the enol species, **14**. Subsequent ketonization of **14** produces **8**. It has previously been demonstrated that **8**, in aqueous solution, will form both the hydrate, **12,** and the enol, **13** (26), accounting for the presence of these products in the spectrum. Interestingly, **7** is a reasonably good substrate for CaaD with a $k_{cat}/K_m$ value nearing that measured for **1** and **11**.

The observation that **7** is a substrate and not an inhibitor reflects the difference between the active sites of CaaD and 4-OT. CaaD is set up to do a hydration reaction while 4-OT carries out an isomerization reaction. As a result, the active site of CaaD is likely to be more hydrophilic in nature than the active site of 4-OT. Pro-1 of 4-OT is able to function as a general base at physiological pH because its p$K_a$ is lowered to about 6.4. The hydrophobic active site of 4-OT plays a major role in lowering the p$K_a$ of Pro-1 and in providing a favorable environment for catalysis (10).

The other two acetylene compounds, **9** and **10**, are also substrates for the enzyme, but upon hydration they are converted into potent inhibitors of CaaD, with **9** being shown to modify Pro-1 of the β-subunit. Modification of β–Pro-1 can occur by a mechanism-based route to form an acyl halide (or a ketene) or by Michael addition of β-Pro-1 to C-3 (Scheme 17). In the first mechanism (Scheme 17A), the enzyme

hydrates **9** (or **10**) to generate an unstable species **15**, which will readily undergo a rearrangement to either the acyl halide (**16**), or the ketene (**17**). Due to their reactivity, both species would be lethal to the enzyme.

In the second mechanism (Scheme 17B), Pro-1 attacks the C-3 position of **9** in a Michael-type reaction. While conjugate additions to α,β-unsaturated carboxylic acids are not normally known, the inhibitor would presumably be activated by an interaction between the carboxylate group of **9** and a residue on the enzyme (i.e., α-Arg-11). Subsequent rearrangement would result in the covalent attachment of the propargyl moiety to the enzyme and the loss of the halide ion.

The observation that the covalent adduct on Pro-1 has a molecular mass of 85 Da is consistent with the attachment of a 3-oxoacetate moiety. Moreover, this adduct, a β-ketoacid, is quite susceptible to facile decarboxylation upon covalent attachment, resulting in an acetyl group (37). At first glance, this sequence of events seems consistent with the mechanism-based route. However, the 3-oxoacetate moiety could also be derived from a rearrangement of the propargyl moiety as shown in Scheme 18. A non-enzymatic rearrangement of the propargyl moiety would result in the formation of an electrophilic allene-type compound (**18**), which would be highly vulnerable to attack by a water molecule. Ketonization of the resulting enol could generate the observed 3-oxoacetate moiety.

The detection of the acyl halide, the ketene, or their hydrolysis products, malonate, would provide a key piece of evidence supporting the mechanism-based route to inactivation. Thus far, they have remained elusive and are likely to remain so. If either species is generated by the enzyme, it must immediately alkylate the enzyme as there is no evidence for its release into solution. This premise is supported the observation that the inactivation of CaaD and the bromide ion release upon inactivation are both nearly stoichiometric. If immediate alkylation did not occur, then CaaD would continue to process **9**, and the concentrations of bromide ion release would be greater than the concentrations of enzyme.

While our current data are consistent with either route, the mechanism-based route is favored by the observation that **7** is converted to acetopyruvate (**8**) and recent experiments indicating that β–Pro-1 has a p$K_a$ of ~9.2[4], making it not sufficiently nucleophilic to add to **9** or **10** in a Michael type reaction. Interestingly, if Pro-1 functions as a general acid catalyst in the reaction, the initial hydration of **9** (or **10**) by CaaD would remove a proton from Pro-1 and render it nucleophilic and quite susceptible to acylation by the acyl halide or reaction with the ketene. Further studies are being actively pursued to discriminate between these mechanisms and a crystal structure of the inactivated enzyme will soon be available.

Our studies with the three acetylene compounds raise an intriguing possibility about the biological function of CaaD as well as its evolution. The CaaD-catalyzed processing of the acetylene compounds is consistent with a hydratase reaction, suggesting that this might be the true physiological role of CaaD in the organism (or perhaps the original role). The loss of the halide from **1** or **11**, after the addition of water, might be a fortuitous side reaction resulting from the very rapid non-enzymatic decay of the unstable halohydrin. This would make CaaD an "accidental" dehalogenase (28), which might have been conscripted to serve in the catabolic pathway for 1, 3-dichloropropene. This speculation is further supported by the observation that CaaD does not appear to have a halide-binding pocket (3). Identification of the putative physiological substrate for CaaD would further assist in the delineation of its evolution.

Whatever the original function of CaaD is in the organism, the relationships between its α- and β-subunits and 4-OT and YwhB are interesting and raise questions about their evolutionary lineage. In one possible scenario, an ancestral chromosomally encoded enzyme, having a non-specific tautomerase activity as well as other low-level activities, could give rise to the 4-OT activity and the CaaD activity through a duplication event followed by a series of mutations to obtain the desired activity (38, 39). Both 4-OT and CaaD retain the ability to carry out a common chemical step, which is the protonation of the carbon adjacent to the

carboxylate group, involving Pro-1 and Arg-11. In 4-OT, the Arg-11 functions as an electron sink to draw electron density toward C-5 for protonation by Pro-1. In CaaD, Arg-11 may also function as an electron sink to draw electron density away from C-3 (resulting in a partial positive charge at C-3) and thereby facilitate protonation at C-2 by Pro-1. While the physiological function of YwhB is not yet known, we have recently determined that both 4-OT and YwhB have low-level CaaD activities, which may result from the presence of this catalytic core (7). These observations reinforce the hypothesis that the β–α–β structural motif is a versatile template (7), which Nature has clearly exploited to create the new activity and structure displayed by CaaD.

**ACKNOWLEDGMENTS**

Electrospray ionization (ESI) and matrix assisted laser desorption-ionization (MALDI) mass spectrometry was performed by the Analytical Instrumentation Facility Core (College of Pharmacy, The University of Texas at Austin) supported by Center grant ES 07784. We thank Steve D. Sorey (Department of Chemistry, The University of Texas at Austin) for his expert assistance in acquiring the NMR spectra, Dr. Maria D. Person (Division of Pharmacology and Toxicology, The University of Texas at Austin) for obtaining and analyzing mass spectral data, and Dr. William H. Johnson, Jr. (Division of Medicinal Chemistry, The University of Texas at Austin) for synthesizing needed compounds and interpreting the NMR spectra.

**REFERENCES**


1. Hartmans, S., Jansen, M. W., van der Werf, M. J., and de Bont, J. A. M. *J. Gen. Microbiol.,* 1991, **137**, 2025-2032.
2. Van Hylckama Vlieg, J. E. T., and Janssen, D. B. *Biodegradation,* 1992, **2**, 139-150.
3. Poelarends, G. J., Saunier, R., and Janssen, D. B. *J. Bacteriol.,* 2001, **183**, 4269-4277.
4. Poelarends, G. J., Kulakov, L. A., Larkin, M. J., van Hylckama Vlieg, J. E. T., and Janssen, D. B. *J. Bacteriol.,* 2000, **182**, 2191-2199.
5. Poelarends, G. J., Wilkens, M., Larkin, M. J., van Elsas, J. D., and Janssen, D. B. *Appl. Environ. Microbiol.,* 1998, **64**, 2931-2936.
6. Subramanya, H. S., *et al. Biochemistry*, 1996, **35**, 792-802.
7. Whitman, C. P. *Arch. Biochem. Biophys.,* **2002, 402**, 1-13.
8. Harris, T. K., *et al. Biochemistry,* 1999, **38**, 12343-12357.
9. Czerwinski, R. M., *et al. Biochemistry,* 1999, **38**, 12358-12366.
10. Czerwinski, R. M., Harris, T. K., Massiah, M. A., Mildvan, A. S., and Whitman, C. P. *Biochemistry,* 2001, **40**, 1984-1995.
11. Marletta, M. A., Cheung, Y.-F., and Walsh, C. T. *Biochemistry,* 1982, **21**, 2637-2644.
12. Johnson, Jr., W. H., Czerwinski, R. M., Fitzgerald, M. C., and Whitman, C. P. *Biochemistry,* 1997, **36**, 15724-15732.
13. Strauss, F., Kollek, L., and Heyn, W. *Chem. Ber,* 1930, **63**, 1868-1899.
14. Andersson, K. *Chem. Scripta,* 1972, **2**, 117-120.
15. Sambrook, J., Fritsch, E. F., & Maniatis, T. *Molecular Cloning: A Laboratory Manual,* 1989.
16. Laemmli, U. K. *Nature,* 1970, **227**, 680-685.
17. Waddell, W. J. *J. Lab. Clin. Med.,* 1956, **48**, 311-314.

18. Ye, Q.-Z., Johnson, L. L., and Baragi, V. *Biochem. Biophys. Res.Commun.,* 1992, **186**, 143-149.
19. Holler, T. P., Foltin, S. K., Ye, Q.-Z., and Hupe, D. J. *Gene,* 1993, **136**, 323-328.
20. Shine, N. R., *et al. Bioorganic Chem.,* 2002, **30**, 249-263.
21. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. *Science,* 2001, **291**, 1790-1792.
22. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. *Gene,* 1989, **77**, 51-59.
23. Marcotte, P., and Walsh, C.T. *Biochemistry*, 1978, **17**, 5613-5619.
24. Gil, A. M., *et al. J. Agric. Food Chem.,* 2000, **48**, 1524-1536.
25. Guthrie, J. P. *JACS,* 1972, **94**, 7020-7024.
26. Stamps, S. L., Fitzgerald, M. C., and Whitman, C. P. *Biochemistry,* 1998, **37**, 10195-10202.
27. Person, M. D., Monks, T. J., and Lau, S. S. *Chem. Res. Toxicol*., 2003 (in press).
28. Copley, S. D., in *Comprehensive Natural Products Chemistry,* Vol. 5, pp 401-422, 1999.
29. Verschueren, K. H. G., Seljée, F., Rozeboom, H. J., Kalk, K. H., and Dijsktra, B. W. *Nature,* 1993, **363**, 693-698.
30. Newman, J., *et al. Biochemistry,* 1999, **38**, 16105-16114.
31. Yang, G., *et al. Biochemistry,* 1996, **35**, 10879-10885.
32. Murzin, A. G. *Curr. Opin. Struct. Biol.,* 1996, **6**, 386-394.
33. Almrud, J. J., *et al. Biochemistry,* 2002, **41**, 12010-12024.
34. Czerwinski, R. M., *et al. Biochemistry,* 1997, **36**, 14551-14560.
35. Braddon, S. A., and Dence, C. W. *Tappi,* 1968, **51**, 249-256.
36. Kosower, E. M. Physical Organic Chemistry, 1968, p 81.
37. Jencks, W. P. *Catalysis in Chemistry and Enzymology*, 1987, pp 116-120.
38. Palmer, D. R. J., *et al. Biochemistry,* 1999, **38**, 4252-4258.

39. O'Brien, P. J., and Herschlag, D. *Chem. Biol.,* 1999, **6**, R91-R105.

**FOOTNOTES**

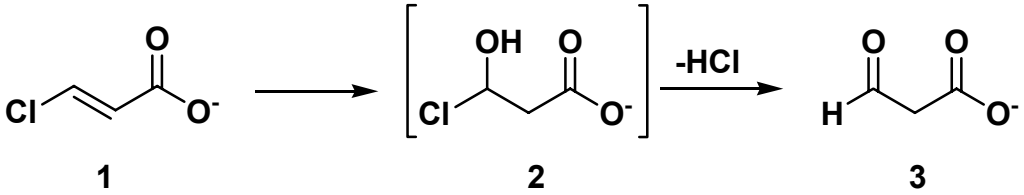
[1]Abbreviations: Ap, ampicillin; CaaD, *trans*-3-chloroacrylic acid dehalogenase; Cm, chloramphenicol; DMSO, dimethyl sulfoxide; ESI-MS, electrospray ionization mass spectrometry; HPLC, high pressure liquid chromatography; IPTG, isopropyl-β-D-thiogalactoside; Kn, kanamycin; LB, Luria-Bertani Medium; MALDI-PSD, matrix assisted laser desorption-ionization post-source decay; MALDI-TOF, matrix assisted laser desorption-ionization time-of-flight; NMR, nuclear magnetic resonance; 4-OT, 4-oxalocrotonate tautomerase; PCR, polymerase chain reaction; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TFA, trifluoroacetic acid

[2]Trypsin has cleaved at the C-terminal side of Leu-11. We attribute this to the presence of chymotrypsin in the non-TPCK-treated trypsin.

[3]While the parent peptides selected for fragmentation have masses of 1242 and 1244, the b ions produced show only a mass addition of +42, instead of the expected additions of +42 and +44. One possibility is that a gas-phase rearrangement of the fragments occurred to produce only the +42 fragments.

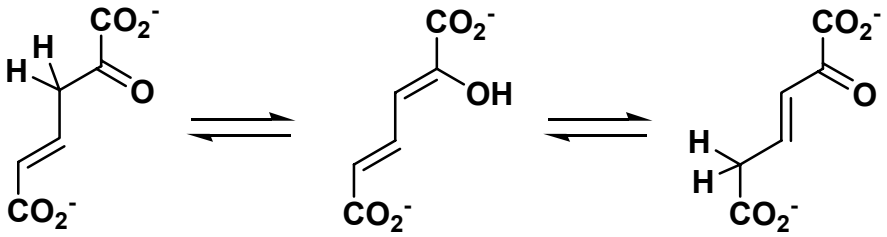
[4]H.F. Azurmendi, S.C. Wang, M.A. Massiah, C.P. Whitman, A.S. Mildvan, (2003) unpublished results.

**Scheme 12**



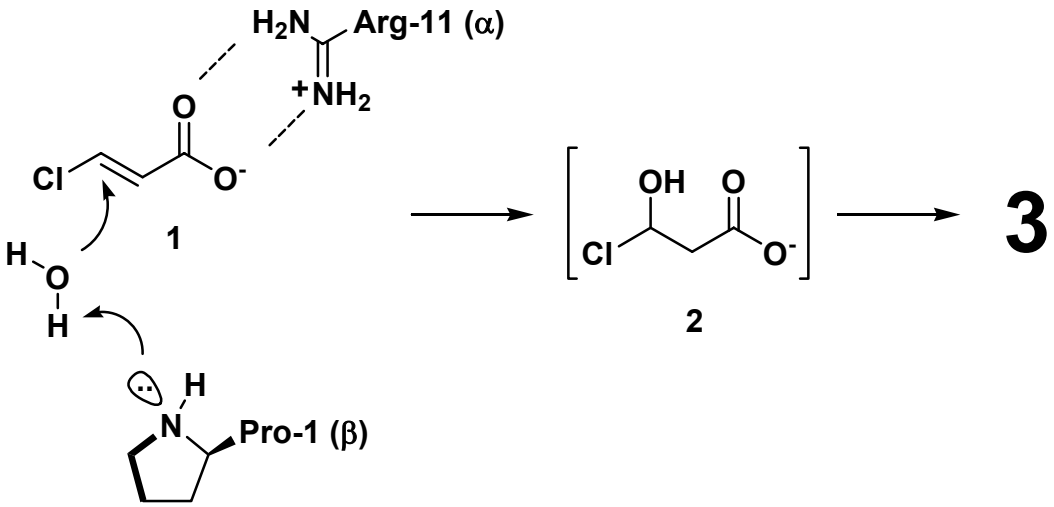

1 2 3

**Scheme 13**

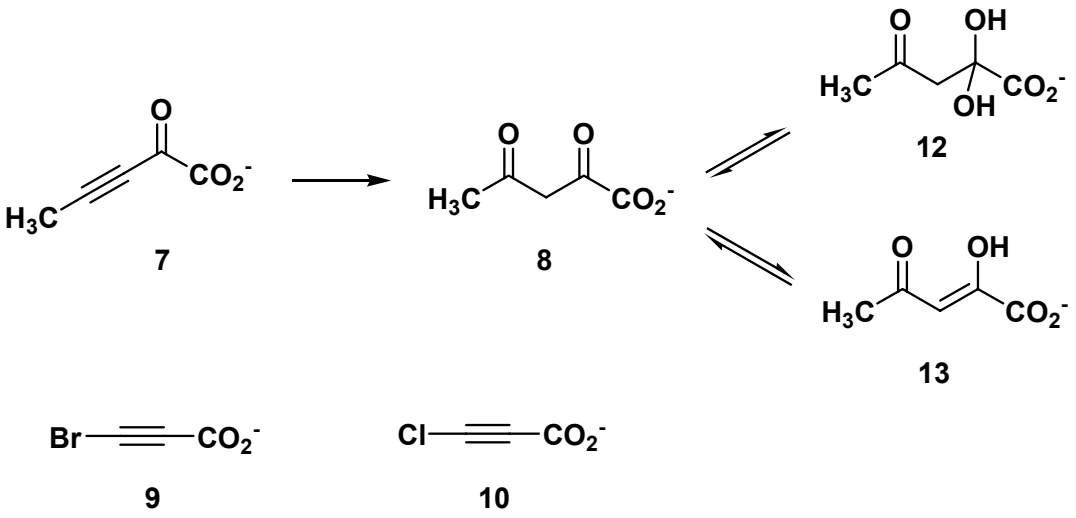



**4** **5** **6**

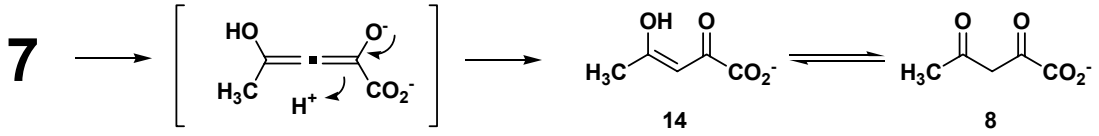
**Scheme 14**



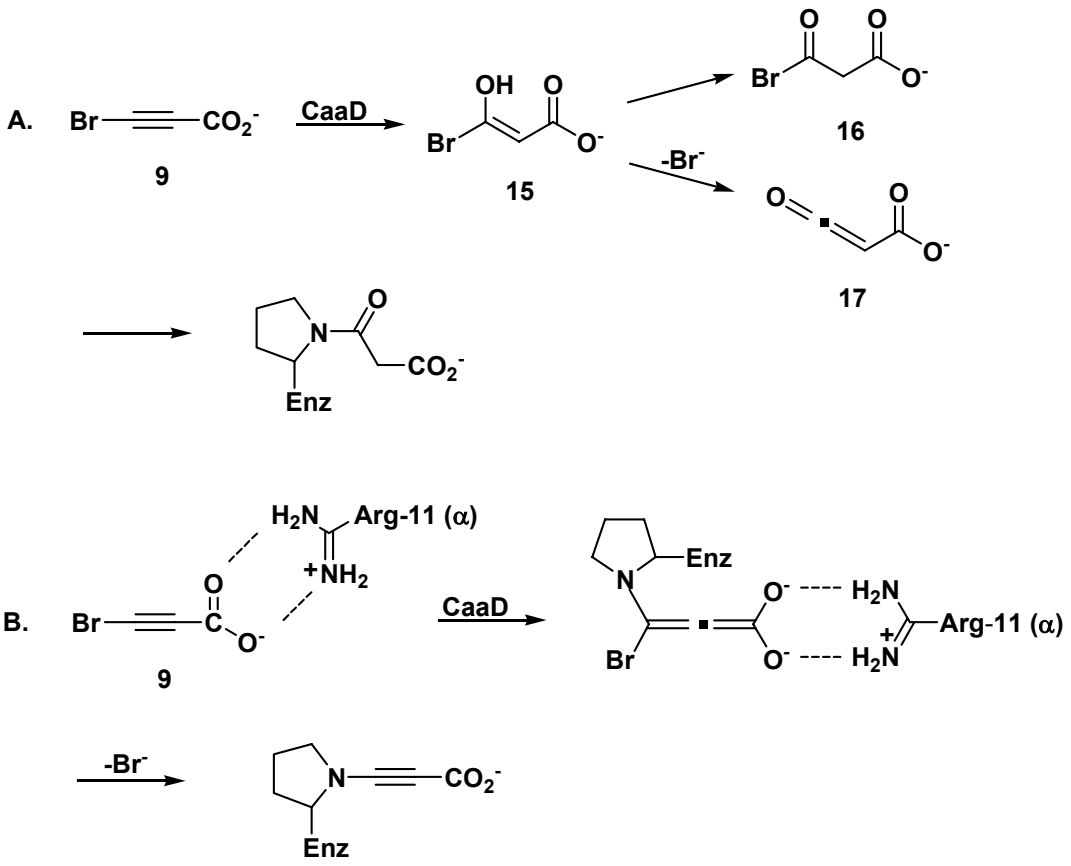
**Scheme 15**


O
CO2-
H3C
7
O
O
H3C
CO2-
8
O
OH
H3C
CO2-
OH
12
O
OH
H3C
CO2-
13
Br
CO2-
9
Cl
CO2-
10

**Scheme 16**



**Scheme 17**


A.
Br
CO2-
9
CaaD
OH
O
Br
O-
15
O
O
Br
O-
16
-Br-
O
O
O-
17
N
O
CO2-
Enz
B.
H2N
Arg-11 (α)
+NH2
O
Br
C
O-
9
CaaD
Enz
N
O-
H2N
Arg-11 (α)
Br
O-
H2N
-Br-
N
CO2-
Enz

**Scheme 18**

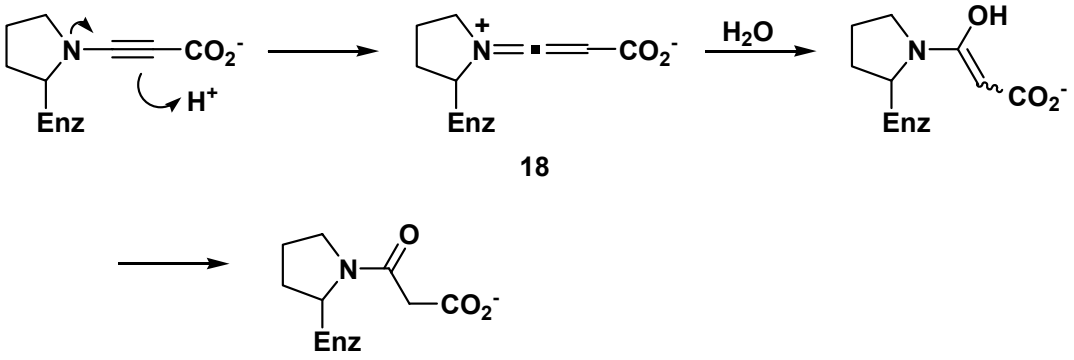


**Figure 29**. $^{1}H$ NMR (500-MHz, $H_2O$) spectra monitoring the CaaD-catalyzed conversion of **1** to **3**. A. The initial spectrum is acquired 5 min after the addition of CaaD to a NMR tube containing **1**. The signals corresponding to protons on **1** and **3** are labeled. The signals at 2.34 ppm and 5.15 ppm correspond to the hydrate of **3**. B. The spectrum acquired 30 min after the addition of enzyme. The signal at 2.0 ppm corresponds to the methyl group of acetaldehyde (*19*).

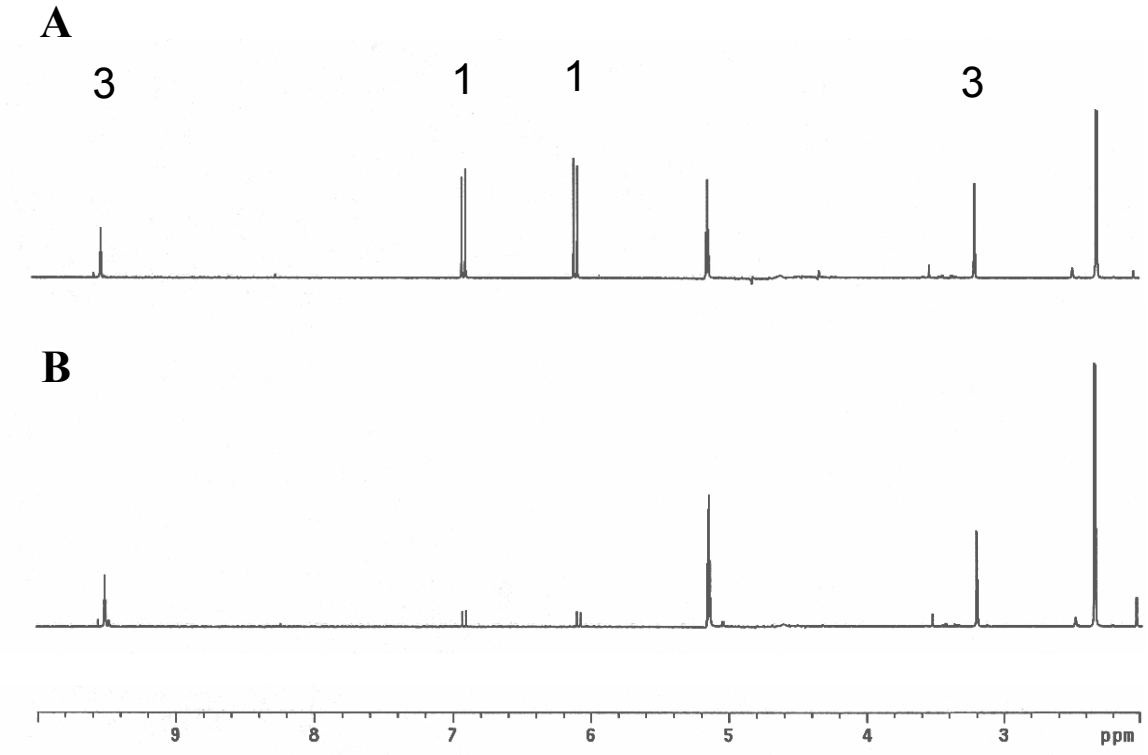


**Figure 30**. UV spectra (1-mm path length) following the CaaD-catalyzed conversion of **7** to **8**. Spectra were recorded every 30 s until the reaction neared completion (~15 min). The decrease in absorbance at 234 nm corresponds to the loss of **7** (*12*), while the production of **8** corresponds to an increase in absorbance at 294 nm (*11*).

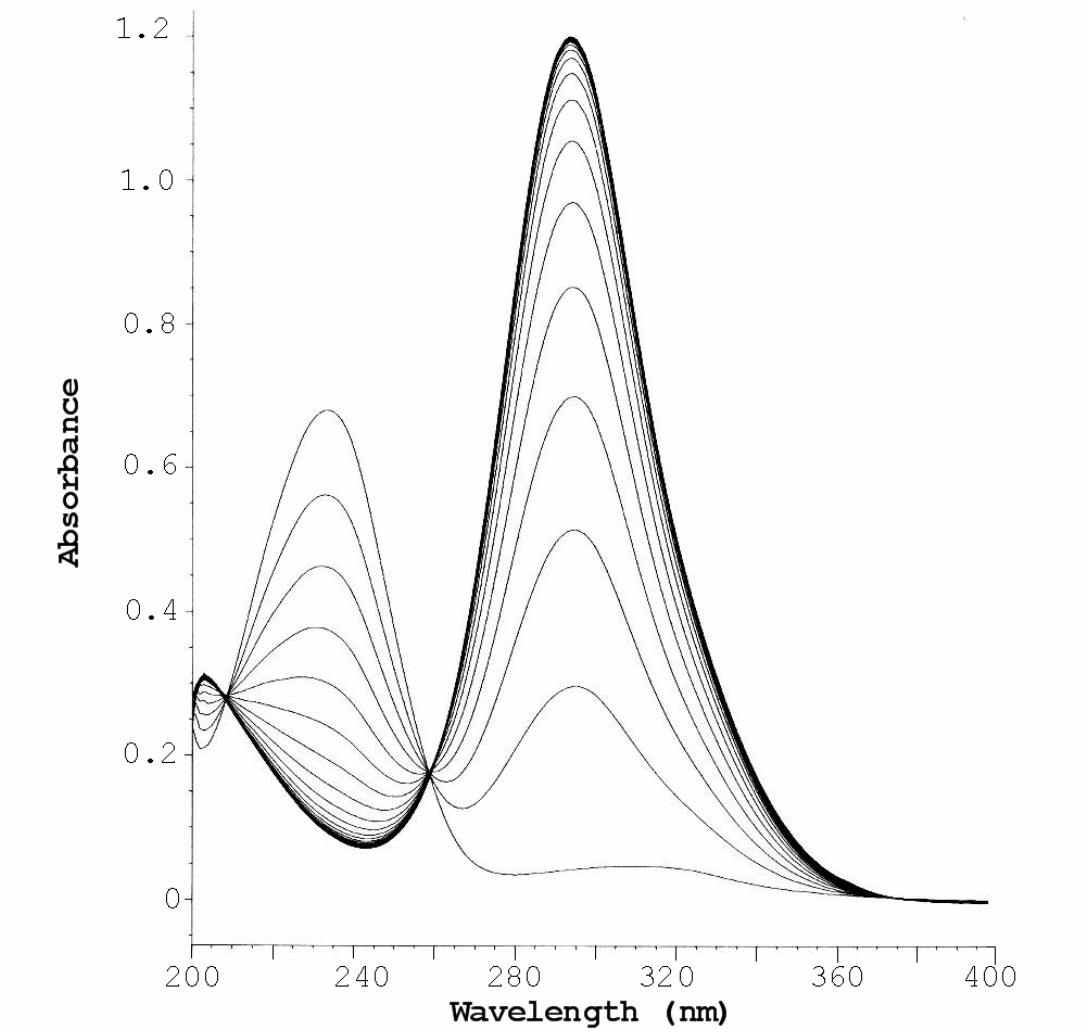


**Figure 31**. The inactivation of CaaD after incubation with varying amounts of **9** or **10** (filled diamonds, 0 μM; filled squares, 20 μM; filled triangles, 40 μM). The concentration of CaaD used is these experiments was 36 μM. The inactivation of CaaD by **9** is shown in panel A while the inactivation of CaaD by **10** is shown in panel B.

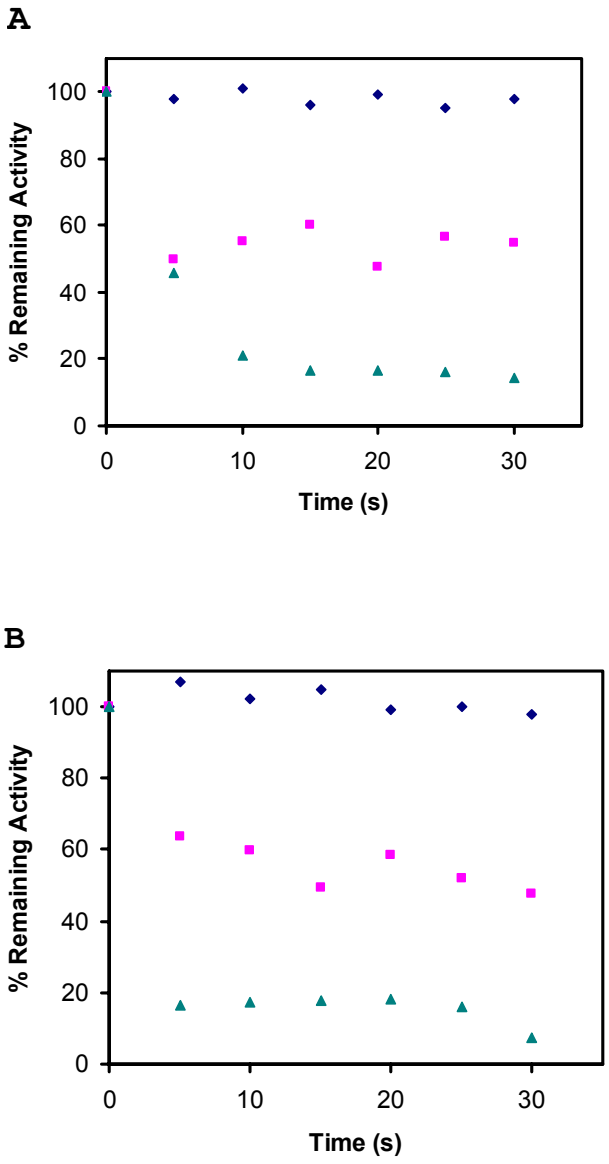


**Figure 32**. Analysis of bromide ion release from various amounts of **9** (filled diamonds, 0 μM; filled squares, 25 μM; filled triangles, 150 μM; crosses, 300 μM) upon incubation with a fixed amount of CaaD (127 μM).

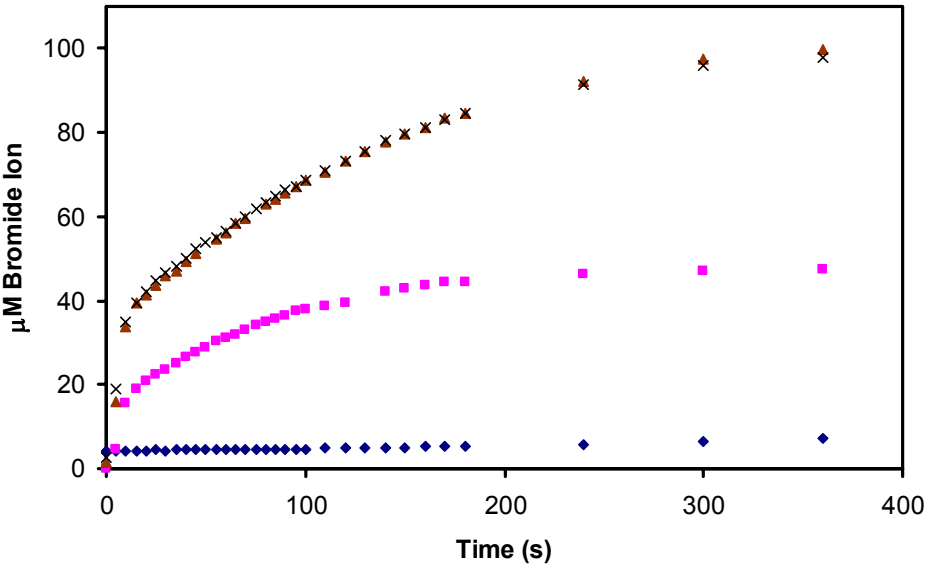
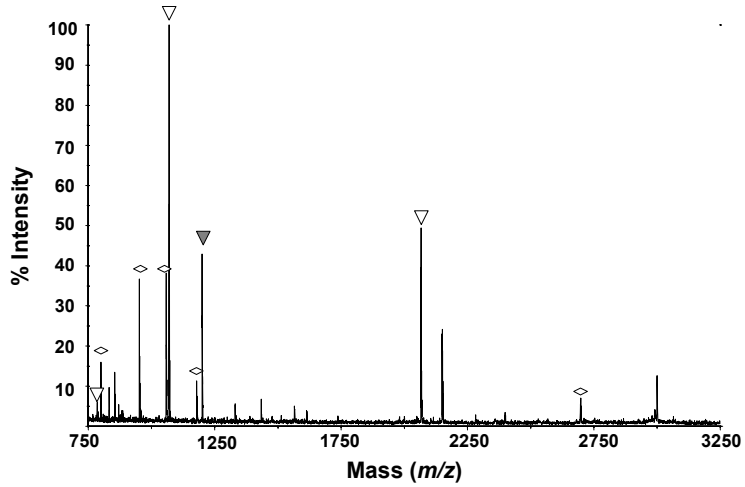


**Figure 33**. MALDI-TOF mass spectral analysis of the tryptic digest of A) a control sample containing CaaD and B) a CaaD sample treated with **9**. The signals corresponding to fragments of the β-subunit are labeled with open triangles while the filled triangle designates the signal at 1200.63 ± 0.06 Da (in panel A) and at 1242.68 ± 0.06 and 1244.75 ± 0.12 Da (in panel B). The calculated and observed masses for these signals are summarized in Table 4. The signals corresponding to fragments of the α-subunit are labeled with open diamonds.

**A**




**B**

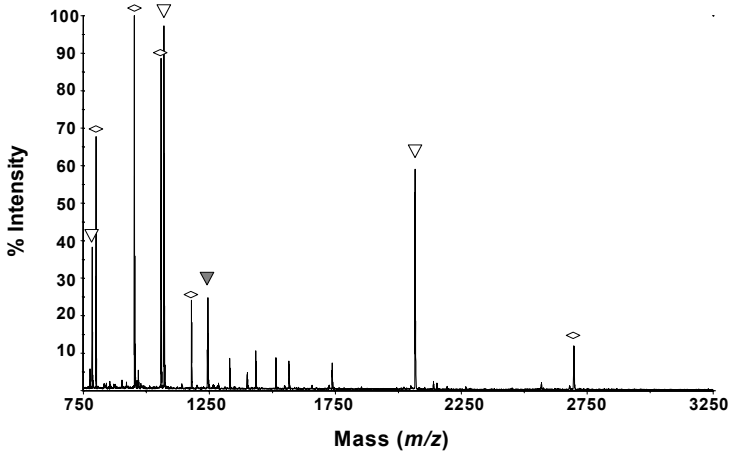


**Figure 34**. MALDI-PSD fragmentation spectra of the peaks at A) 1200.63 Da in the control digest, corresponding to the PFIECHIATGL (1-11) fragment of the β-subunit of CaaD and at B) 1242.68 and 1244.75 in the treated digest, corresponding to the PFIECHIATGL (1-11) modified by adducts of 42 and 44 Da, respectively. The b ions and the $b_1$ ions (in panels A and B, respectively) are labeled with open triangles. The $b_1$ ions are shifted by +42 Da.

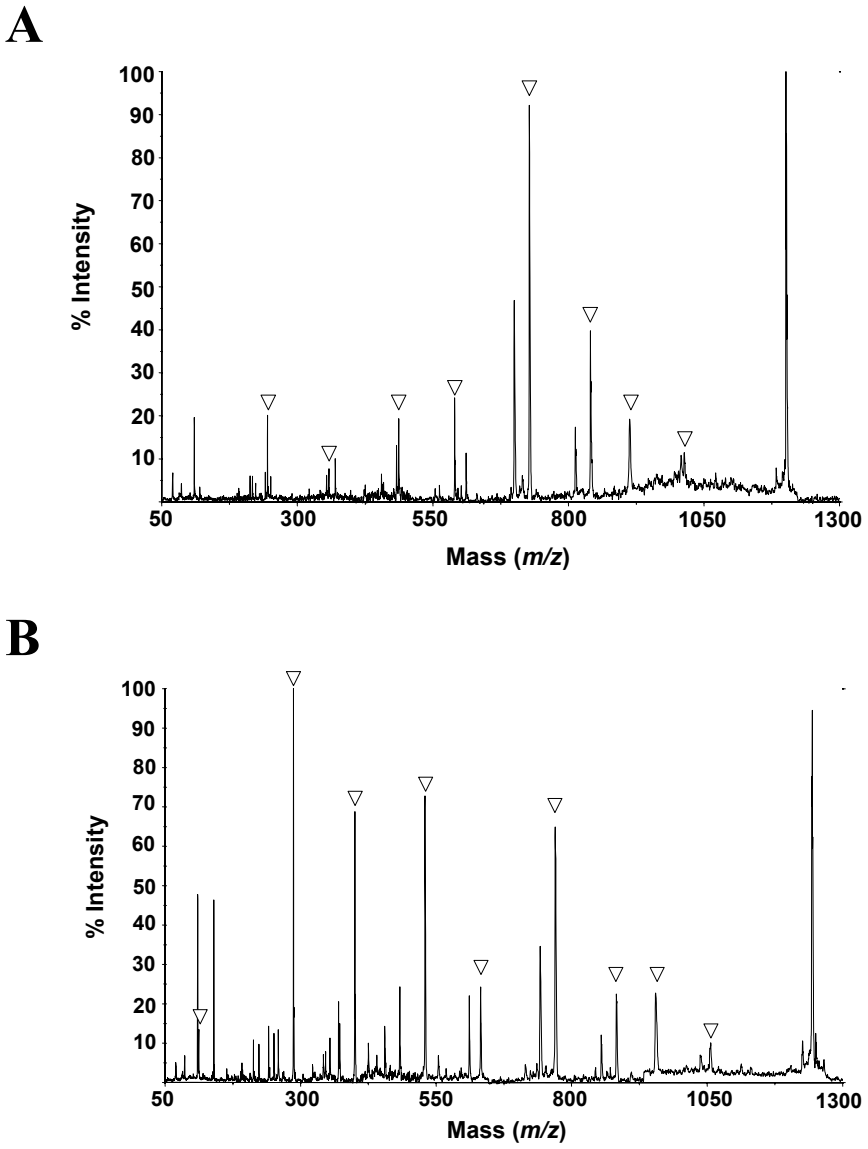


**Table 2. Kinetic Parameters for CaaD[a]**

| substrate | $k_{cat}$ ($s^{-1}$) | $K_m$ ($\mu M$) | $k_{cat}/K_m$ ($M^{-1}s^{-1}$) |
|---|---|---|---|
| **1** | 3.8 ± 0.1 | 31 ± 2 | $1.2 \times 10^5$ |
| **11** | 5.1 ± 0.1 | 37 ± 2 | $1.4 \times 10^5$ |
| **7** | 0.7 ± 0.02 | 110 ± 4 | $6.4 \times 10^3$ |

[a]The steady-state kinetic parameters were determined in 20 mM sodium phosphate buffer (pH 9.0) at 23 °C. Errors are standard deviations.

**Table 3. Protection of CaaD by 11.**

| **[I]** | 1 mM **11** | 5 mM **11** | 10 mM **11** |
|---|---|---|---|
| 20 μM **9**[a] | 35% (10 s)[b] | 35% (40 s) | 70% (70 s) |
| 40 μM **9** | 40% (10 s) | 50% (30 s) | 40% (60 s) |
| 20 μM **10** | 65% (10 s) | 40% (35 s) | 25 % (150 s) |
| 40 μM **10** | 40 % (10 s) | 45% (40 s) | 20% (130 s) |

[a]CaaD was incubated with the indicated concentrations of **9** or **10** in 20 mM sodium phosphate buffer (pH 7.3) at 23 °C along with the indicated concentrations of **11**.

[b]The time interval indicates the approximate amount of activity remaining after the indicated time.

**Table 4. Mass Spectral Analysis of the Tryptic Digest of the β-Subunit.**

| Peptide fragment | Calculated mass[a] | Observed mass (control) | Observed mass (treated) |
|---|---|---|---|
| 1-11 | 1200.61 | 1200.63 | 1242.68/1244.75 |
| 16-21 | 785.50 | 785.51 | 785.53 |
| 22-29 | 903.48 | Not detected | 903.53 (weak) |
| 30-36 | 703.36 | Not detected | 703.39 (weak) |
| 37-55 | 2066.11 | 2066.16 | 2066.16 |
| 56-65 | 1070.52 | 1070.54 | 1070.58 |
| 66-70 | 474.26 | Not detected | Not detected |

[a]These values were calculated using the monoisotopic $MH^+$ masses (Da).

## pH RATE ANALYSES OF WILD-TYPE AND α-P1A *TRANS*-3-CHLOROACRYLIC ACID DEHALOGENASE

### INTRODUCTION

The recently discovered alkene dehalogenase, *trans*-3-chloroacrylic acid dehalogenase (CaaD), converts the *trans* isomers of 3-haloacrylates (**1** or **2**, Scheme 19) to the unstable halohydrin intermediate (**3**) by the addition of water. Subsequent collapse of **3** affords malonate semialdehyde (**4**) (1). The $(\alpha\beta)_3$ heterohexameric CaaD is elaborated by the soil bacterium *Pseudomonas pavonaceae* 170 as part of a degradative pathway for 1,3-dichloropropene, the active ingredient in the commercially available fumigants, Shell D-D and Telone II (2,3). These compounds are mixed in with the soil to kill nematode parasites. Sequence analysis placed CaaD in the 4-oxalocrotonate tautomerase (4-OT) family of the tautomerase superfamily (1). Conserved residues include the N-terminal proline, found in both subunits, and an arginine at position 11, found only in the α-subunit. Previous site-directed mutagenesis experiments identified Pro-1 of the β-subunit and Arg-11 of the α-subunit as critical mechanistic residues (1). These experiments have also suggested the Pro-1 of the α-subunit is not critical for catalysis, since the observed activity of the α-P1A mutant was comparable with those of the wild-type enzyme (1).

Tautomerase superfamily members are structurally homologous proteins, which share a common β–α–β structural fold as well as a catalytic Pro-1 (4, 5). In the bacterial isomerase 4-OT, the best-characterized enzyme in the superfamily, Pro-1 functions as a general base. The overall reaction transforms β,γ–unsaturated enones to their α,β-isomers via a dienol intermediate (4). Pro-1 of 4-OT can function effectively as a general base at cellular pH because it has a p$K_a$ of ~ 6.4, as determined by a pH rate profile and direct NMR titration (6, 7). The p$K_a$ is lowered ~3 pH units from that observed for the proline in the model compound, proline amide, and is largely the result of its presence in a hydrophobic active site. Phe-50 plays a major role in maintaining the hydrophobic environment (8). An amino acid with an

aromatic or hydrophobic side chain is conserved in this position in most known 4-OT family members. This observation and the fact that that Pro-1 has been implicated as the catalytic base in superfamily members characterized to date, suggested that Pro-1 might behave similarly in homologues such as CaaD.

On the basis of these observations, two potential roles have been proposed for the essential β-Pro-1 (Scheme 20) (1). In one mechanism, Pro-1 is a general base that activates water for attack at C-3 of **1**. The proton at C-2 comes from solvent or Pro-1, now functioning as a general acid catalyst. In the second mechanism, an unknown residue (i.e., B:, the general base) activates the water molecule for attack at C-3 while the proton at C-2 comes from Pro-1, a general acid catalyst. To discriminate between these two mechanisms, we determined the p$K_a$ of β-Pro-1 by pH rate studies. If the p$K_a$ is substantially lowered, as it is in 4-OT, it could be inferred that β-Pro-1 acts as a general base. If the p$K_a$ is comparable to that of proline amide (p$K_a$ ~ 9.4), β-Pro-1 cannot function as a general base at cellular pH.

To gain insight into the precise roles that the N-terminal prolines play in the catalytic mechanism of CaaD, we first conducted a pH rate analysis of the wild-type enzyme by UV spectroscopy using *trans*-3-bromoacrylate (**1**) as the substrate. Site-directed mutagenesis was used to construct the α-P1A-CaaD mutant, which was over-expressed and purified to near homogeneity. The kinetic parameters for the α-P1A mutant were determined under the conditions used for the wild-type enzyme. As has been shown elsewhere, the α-Pro-1 did not appear to be important for catalysis because its kinetic parameters were similar to those of the wild-type enzyme. A kinetic pH rate analysis of the mutant enzyme was also performed to distinguish the role of α-Pro-1. Both the wild-type and α-P1A-CaaD show a bell-shaped dependence on pH, confirming the requirements for a basic and an acidic group in the mechanism.

These results indicate that Pro-1 of the α-subunit does not play a role as either an acid or a base catalyst. Moreover, a substantial decrease is not observed for any of the p$K_a$ values, indicating that β-Pro-1 does not act as a general base. Both the acidic and basic groups are attributable to the enzyme. The acidic carboxylate group of

*trans*-3-bromoacrylate can be ruled out because its p$K_a$ value was measured at ~4.6. To assign the observed p$K_a$ values to specific peaks, literature procedures were modified to provide uniformly $^{15}$N-labeled protein for $^{15}$N-NMR spectroscopy (9, 10). A brief discussion of the NMR titration results, which were performed in collaboration with Dr. Albert S. Mildvan (Johns Hopkins University), is described here. The NMR data assigns a p$K_a$ of ~9 to β-Pro-1, suggesting that it acts as an acid catalyst. These results, taken together, identify a new role for the conserved amino-terminal proline in the tautomerase superfamily (4, 5).

## MATERIALS AND METHODS

*Materials.* All reagents, buffers, and solvents were obtained from Aldrich Chemical Co. (Milwaukee, WI), Sigma Chemical Co. (St. Louis, MO), Fisher Scientific Inc. (Pittsburgh, PA), or Spectrum Laboratory Products, Inc (New Brunswick, NJ) with the following exceptions. The *trans*-3-chloroacrylate (**2**) was obtained from Fluka Chemical Corp. (Milwaukee, WI). Literature procedures were used for the syntheses of the *trans*-isomer of 3-bromoacrylate (**1**) (11) and 2-oxo-3-pentynoate (**5**) (12) and the purification of CaaD (13). Tryptone, yeast extract, and agar were obtained from Becton, Dickerson, and Company (Franklin Lakes, NJ). The Ultrafree DA centrifugal filter units, the YM-3 ultrafiltration membranes, and the Amicon concentrators were obtained from Millipore Corp. (Billerica, MA). The thin-walled PCR tubes were obtained from Ambion, Inc. (Austin, TX). Restriction enzymes, PCR reagents, and T4 DNA ligase were obtained from Promega Corp. (Madison, WI) or New England Biolabs (Beverly, MA). The Wizard PCR Preps DNA purification systems were purchased from Promega Corp. The QIAprep Spin Miniprep kit was obtained from Qiagen, Inc. (Valencia, CA). Herculase DNA Polymerase was obtained from Stratagene (La Jolla, CA). The remaining molecular biology reagents including agarose, DNA ladders, and protein molecular weight standards, were obtained from Invitrogen Corporation (Carlsbad, CA). Oligonucleotides for DNA amplification and sequencing were synthesized by either Oligos Etc. (Wilsonville, OR) or Genosys (The Woodlands, TX).

*Strains and Plasmids. Escherichia coli* strains DH5α (Invitrogen) and JM109 (Promega) were used for cloning and isolation of plasmids. *E. coli* strains BL21(DE3)pLysS from Novagen and BL21-Gold(DE3)pLysS from Stratagene (La Jolla, CA) were used for recombinant protein expression. The plasmids containing the gene for the α-subunit of CaaD [pET-24a(+)-CaaDα] and the gene for the β-subunit of CaaD [pET-21a(+)-CaaDβ4] were constructed as described elsewhere (13).

*Methods*. Techniques for restriction enzyme digestions, ligation, transformation, and other standard molecular biology manipulations were based on methods described elsewhere (8), or as suggested by the manufacturer. Plasmid DNA was introduced into cells by electroporation using a Cell-Porator Electroporation System (Whatman Biometra, Göttingen, Germany). The PCR was performed in a DNA thermal cycler (Model 480) obtained from PerkinElmer Inc (Wellesley, MA). DNA sequencing was done at the DNA Core Facility in the Institute for Cellular and Molecular Biology at the University of Texas (Austin). Kinetic data were obtained on an Agilent 8453 Diode Array spectrophotometer. Solutions in the cuvettes were mixed using a stir/add cuvette mixer (Bel-Art Products, Pequannock, NJ). The kinetic data were fitted by nonlinear regression data analysis using the Grafit program (Erithacus Software Ltd., Horley, U. K.) obtained from Sigma Chemical Co. HPLC was performed on a Waters (Milford, MA) 501/510 system or a Beckman System Gold HPLC (Fullerton, CA) using a TSKgel Phenyl-5PW hydrophobic column (Tosoh Bioscience, Montgomeryville, PA). Protein was analyzed by Tris glycine sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions on 17.5% gels using either the Mini-Protean II or III vertical gel electrophoresis apparatus obtained from Bio-Rad (Hercules, CA) (14). Protein concentrations were determined using the method of Waddell (15).

*Site-Directed Mutagenesis*. The α-P1A-CaaD mutant was constructed using the gene for the α-subunit of CaaD [in pET-24a(+)] as the template (13). The gene is flanked by an *Nde*I restriction site and a *Not*I restriction site. The mutation was introduced by overlap extension PCR (16). The external PCR primers were oligonucleotides 5'-GCG GAT AAC AAT TCC CCT CT-3' (designated primer A) and 5'-CTC AGC TTC CTT TCG GGC TT-3' (designated primer D). Primer A corresponds to a region 36 bases upstream of the *Nde*I restriction site in the pET-24a(+) vector. Primer D corresponds to the complementary sequence of a region 20 bases downstream of the His-tag region in pET-24a(+). The internal primers were 5'-AGA GAT CAT CGC CAT ATG TAT ATC-3' (primer B) and 5'-GAT ATA CAT

ATG GCG ATG ATC TCT-3' (primer C). In each internal primer, the codon for the desired mutation is underlined and the remaining bases correspond to the coding (primer C) or the complementary sequence (primer B).

For the construction of the AB and CD fragments, the PCR reaction mixture contained 10× Herculase reaction buffer (5 μL), 10 mM dNTPs (1 μL), pET-24a(+) containing the gene for the α-subunit of CaaD (5 μL of a 1:500 dilution), 20 μM stock solutions of primers A and B or C and D (2.5 μL each), Herculase (0.5 μL), and sufficient water to make the final volume 50 μL. The PCR protocol consisted of an initial 2-min incubation period at 92 °C, followed by 30 cycles of the PCR where each cycle contained a denaturation step at 92 °C for 30 s, an annealing step at 60 °C for 30 s, and an elongation step at 72 °C for 60 s, and a final 10-min incubation period at 72 °C. Purification of the AB and CD fragments was performed using gel electrophoresis and the Wizard PCR Preps DNA purification system. Construction and purification of the AD fragment was carried out as described above except that the external primers were used, and the AB and CD products (1 μL each) were used as template. Restriction digests of the pET-24a(+) vector and the AD fragment were performed at 37 °C for 6 hours followed by purification using gel electrophoresis and the Ultrafree-DA units. The ligation reaction was performed using the Quick Ligation Kit (New England Biolabs, Inc., Beverly, MA). The ligated plamids were purified by ethanol precipitation and re-suspended in sterile water (5 μL). The purified DNA (2 μL) was used to transform *E. coli* DH5α by electroporation. Transformed cells were plated onto LB/Kn (50 μg/mL) plates for selection. Two colonies were selected at random for insert sequencing. One colony, designated pET-24a(+)-CaaD-αP1A-1, had only the desired mutation.

*Overexpression and Purification of the α-P1A-CaaD Mutant*. For expression, pET-24a(+)-CaaD-αP1A 1 was co-transformed into *E. coli* strain BL21-Gold(DE3)pLysS by electroporation along with pET-21a(+)-CaaDβ4 (containing the gene for the β-subunit of CaaD). Cells were selected in triple antibiotic media (13), and subsequent growth and purification were performed as described for wild-type

CaaD. Typically a 3L culture of the αP1A-CaaD mutant yields 9-10 g of cells and 25-50 mg of protein, which was >95% homogenous as assessed by SDS-PAGE. The observed monomer masses for the α- and β-subunits of αP1A-CaaD were 8315.9 (calc. 8318.0) and 7505.3 (calc. 7507.0), respectively.

*Kinetic Characterization of α-P1A-CaaD.* The kinetic assays were performed as described elsewhere for the wild-type enzyme (13). The kinetic assays were performed in 20 mM $Na_2HPO_4$ buffer, pH 9.0, observing the decrease in absorbance at 224 nm corresponding to the hydration of *trans*-3-bromoacrylate (**1**, $\varepsilon = 9700\ M^{-1}\ cm^{-1}$) and *trans*-3-chloroacrylate (**2**, $\varepsilon = 4900\ M^{-1}\ cm^{-1}$). An aliquot of α-P1A-CaaD (240 μL of a 5.6 mg/mL solution) was diluted into buffer (40 mL) and incubated for 1 h. Subsequently, a 1 mL-portion of the enzyme was transferred to a cuvette and assayed by the addition of a small quantity of substrate (**1** or **2**) from a 5 mM or 50 mM stock solution (in 100 mM $Na_2HPO_4$ buffer, pH 9.2). The addition of **1** or **2** (as their free acids) to the 100 mM phosphate buffer adjusted the pH of the stock solution to about 7.0. The 5 mM stock solution was made by dilution of an aliquot of the 50 mM stock solution into 100 mM $NaH_2PO_4$ buffer, pH 7.3. The concentrations of substrate used in the assay ranged from 10-150 μM.

The hydration of 2-oxo-3-pentynoate (2-OP, **5**, Scheme 21) by CaaD was monitored by following the formation of acetopyruvate (**6**) at 294 nm ($\varepsilon = 7{,}000\ M^{-1}\ cm^{-1}$) in 20 mM sodium phosphate buffer (pH 9.0) (17). An aliquot of α-P1A-CaaD (240 μL of a 2.9 mg/mL solution) was diluted into buffer (50 mL) and incubated for 1 h. The assay was initiated by the addition of a small quantity of **5** from either a 5.4 mM or 53.6 mM stock solution. The stock solution (53.6 mM) was made by dissolving the appropriate amount of 2-oxo-3-pentynoate (8.5 mg) in 100 mM $Na_2HPO_4$ buffer (1.42 mL, pH 9.2). The addition of **5** (as the free acid) adjusted the pH of the buffer to ~7. The 5.4 mM stock solution was made by dilution of an aliquot of the 53.6 mM into 20 mM $NaH_2PO_4$ buffer, pH 7.3. The concentrations of **5** used in the assay ranged from 8-536 μM.

*$^{15}$N-Labeling of Wild Type CaaD and the α-P1A Mutant*. For NMR studies, the wild type and mutant enzymes were $^{15}$N-labeled by the modification of literature protocols (9, 10). The M9 minimal medium was enriched with supplements found in the MOPS minimal media (10) and $^{15}NH_4Cl$ (99% enriched). All other nitrogen sources (e.g., nitrates) were removed. Accordingly, a liter of culture medium contained 5 × M9 salts without $NH_4Cl$ (200 mL), 1M $MgSO_4$ (2 mL), 20% glucose (20 mL), 1 M $CaCl_2$ (100 μL), $B_1$ solution (1.4 mL), vitamin supplement solution (1.4 mL), thymine-uracil solution (1.4 μL), Mg micronutrient solution (1 mL), and 25% $^{15}NH_4Cl$ (2 mL).

The following procedures were used for the growth and purification of the $^{15}$N-labeled enzymes. Cultures of LB media containing both ampicillin (100 μg/mL) and kanamycin (50 μg/mL) were inoculated from single colonies of the appropriate enzyme in its expression strain and grown overnight at 37 °C. Subsequently, cultures (500 mL) of the M9 medium described above and Ap/Kn (100 μg/mL and 50 μg/mL, respectively) were inoculated with 3 mL of the overnight culture and incubated at 30 °C with vigorous shaking. When the $OD_{600}$ reading reached ~0.8, the cultures were induced with IPTG (1 mM) and then allowed to continue to grow at 30 °C. Additional aliquots of ampicillin (50 μg/mL) and kanamycin (25 μg/mL) were added to each culture to maintain the plasmids in the growing cells. After a 24 hr total growth period, cells were harvested by centrifugation and stored at -80 °C until use. Typically 5 L of culture yields 30 g of cells for CaaD, and 25 g for the α–P1A mutant. Due to low yields of enzyme, 10-20 L of culture were processed in each purification procedure, which was performed as described previously. If the protein still contained significant contaminanting protein after the G-200 sizing column (as assessed by SDS-PAGE), the concentrated protein was made 1 M in ammonium sulfate and reloaded onto the phenyl-5PW column for final purification. Typically, the yields per liter of culture are: CaaD, 1-3 mg (> 90% homogenous) and α–P1A-CaaD, 2 mg (>95% homogenous).

*pH Dependence of CaaD Using **1***. The dependence of the rate of hydration of **1** was determined in 20 mM sodium phosphate buffer, pH range 6.5-10.26. Compound **1** was used as the substrate due to its higher extinction coefficient, which greatly facilitated measurements. The buffers were made up by combining appropriate quantities of 20 mM $NaH_2PO_4$, 20 mM $Na_2HPO_4$, and 20 mM $Na_3PO_4$ buffers to maintain constant ionic strength. For each pH value, a sufficient quantity of enzyme (from a stock solution of 8.5 mg/mL in 20 mM $NaH_2PO_4$, pH 7.3) was equilibrated in buffer (40 mL) for ≥ 1 hr at room temperature (29 °C). The addition of the large amount of enzyme changed the pH, so that the reported pH values (6.50-10.26) are those determined after the addition of the enzyme. Subsequently, aliquots (1 mL) were removed and assayed for activity using 12 substrate concentrations ranging from 10-150 μM. Stock solutions of substrate were made up as described (13). The volume of substrate added was 8 μL or less in all experiments. The amounts of enzyme used were 8 μL/mL (pH 6.50), 5 μL/mL (pHs 7.03-7.43), 4 μL/mL (pHs 7.99-8.89), 4.5 μL/mL (pH 8.97), 6 μL/mL (pH 9.44-9.77), 9 μL/mL (pH 10.02), and 10 μL/mL buffer (pH 10.26). In order to show that the activity at high pH was due to the enzymatic reaction (and not the result of denaturation), CaaD (20 μL) was equilibrated at room temperature in 20 mM phosphate buffer (2 mL, pH 10.51) for 2 hrs. An aliquot (200 μL) was subsequently diluted into 20 mM phosphate buffer (pH 9.0) and immediately assayed at two concentrations of (**1**) (100 μM and 150 μM). The resulting activity was comparable to that expected for the wild type at pH 9.0, suggesting that CaaD was not denatured at the high pH values investigated.

*pH Dependence of α-P1A-CaaD Using **1***. The pH rate profile for the α-P1A mutant of CaaD was determined using a procedure similar to that used for the wild-type. Substrate solutions and buffers were made as described for wild-type CaaD. For each pH, a sufficient quantity of enzyme was equilibrated in buffer (40 mL) for ≥ 1 hr at 22 °C. The addition of the enzyme changed the pH, so that the reported pH values are those determined after the addition of the enzyme. The range of final pHs was 6.50-10.82. The amounts of enzyme (18.0 mg/mL stock in 20 mM $NaH_2PO_4$, pH

7.3) used were 1.5 μL/mL (pHs 6.50 and 10.82), 1 μL/mL (pHs 7.06 and 9.50), 0.75 μL/mL (pHs 7.48-9.32), and 2.5 μL/mL (pH 10.26).

*Determination of the $pK_a$ of **1***. A 1 M solution of **1** was made by dissolving the free acid (78.0 mg) in deionized water (517 μL). An equivalent volume of 1 M NaOH was added to the solution, and the pH was measured.

## RESULTS

*Production, Expression, and Characterization of α-P1A-CaaD.* The α-P1A-CaaD mutant was constructed by overlap extension PCR, expressed in *E. coli*, and purified to homogeneity (>95% as assessed by SDS-PAGE) using the procedure for wild-type CaaD described earlier. The sequence of the mutant was confirmed by DNA sequencing to ensure that no other mutations were present. The range of overproduction was 8-17 mg/L. Mass spectrometric analysis confirmed the molecular mass of the mutant α-P1A subunit, and that the N-terminal residues were not blocked by the initiating methionine.

*Kinetic Characterization of α-P1A-CaaD.* The steady-state parameters for the α-P1A-CaaD mutant were determined using **1**, **2**, and **5** as substrates. The observed values for $k_{cat}$, $K_m$, and $k_{cat}/K_m$ are given in Table 5 along with those of the wild-type enzyme (13). These values are similar, confirming that α-P1A is not essential for catalysis. The major difference observed is for the $K_m$ value for **5**, which is 2-fold higher in the wild-type enzyme than the mutant. This is also reflected in the values for $k_{cat}/K_m$, since the $k_{cat}$ values are identical. This may be due to an increase in the opening of the active site or another structural perturbation that allows **5** to bind more readily to the α-P1A mutant.

*Expression and Characterization of Uniformly $^{15}$N-Labeled Enzymes.* Previous efforts for uniform labeling of proteins in the Whitman laboratory have relied on a MOPS-buffered medium (6, 10). To simplify this procedure, a modification of M9 minimal medium was used for the overexpression of labeled CaaD and α-P1A-CaaD. Similar procedures are described elsewhere (18). Due to difficulties with expression, which are likely due to the presence of the $^{15}$N-label compounded by the co-expression of two plasmids, the M9 minimal medium was enriched with several supplements (10). Without supplementation, only low expression of labeled recombinant CaaD was observed.

*pH Dependence of the Kinetic Parameters for CaaD.* The dependence of the rate of hydration of **1** was determined in 20 mM sodium phosphate buffer over the range 6.50 – 10.26. The addition of the large quantity of enzyme significantly altered some pHs, so that the values reported are the final values obtained after mixing buffer with enzyme. A plot of log ($k_{cat}/K_m$) shows a bell-shaped dependence on pH with slopes of 1 on the ascending and descending limbs (Figure 35, open circles). The slopes of unity indicate that a single basic group (ascending limb) and a single acidic group (descending limb) are important for catalysis. These groups are either on the free enzyme or correspond to the carboxylate group of the substrate. A non-linear least-squares fit of the data gives p$K_a$ values of 8.0 ± 0.2 and 9.1 ± 0.2 for these groups. These values can be assigned to the free enzyme since the p$K_a$ of the substrate (**1**) is 4.6 by direct titration.

A plot of log $k_{cat}$ (Figure 35, filled circles) also shows a bell-shaped dependence on pH with slopes of 1 on both the ascending and descending limbs. Similarly, a non-linear least-squares fit for log $k_{cat}$ gives p$K_a$ values of 6.8 ± 0.2 and 9.6 ± 0.2. These results indicate that one acidic and one basic group in the enzyme-substrate complex are important.

*pH Dependence of the Kinetic Parameters for α-P1A-CaaD.* The dependence of the rate of hydration of **1** by the mutant enzyme was determined in 20 mM sodium phosphate buffer over the range 6.50-10.82. Final pH values were those obtained after equilibration of the enzyme in buffer. A plot of log ($k_{cat}/K_m$) shows a bell-shaped dependence on pH with slopes of 1 on the ascending and descending limbs (Figure 36, open circles). The slopes of unity indicate that one basic group (ascending limb) and one acidic group (descending limb), from either the enzyme or the substrate, are important in catalysis. A non-linear least-squares fit of these data gives p$K_a$ values of 7.9 ± 0.2 and 9.0 ± 0.2. These values can be assigned to residues in the free enzyme and not to the substrate.

The plot of log $k_{cat}$ vs. pH (Figure 36, filled circles) shows a similar bell-shaped dependence on pH with slopes of unity on both limbs, indicating that one

basic and one acidic group are involved in the enzyme-substrate complex. The non-linear least-squares fit gives p$K_a$ values of 6.4 ± 0.2 and 9.7 ± 0.2 for the mutant protein. These results indicate that one acidic and one basic group on the mutant enzyme are important in catalysis. Overall, the p$K_a$ values are equivalent for the wild-type and mutant enzymes. This indicates that α-Pro-1 is not involved in catalysis, and that β-Pro-1 does not function as a general base.

*Determination of the p$K_a$ of **1***. The p$K_a$ of **1** was determined by direct titration to be 4.6.

## DISCUSSION

The discovery and initial characterization of *trans*-3-chloroacrylic acid dehalogenase was exciting for a number of reasons. From an evolutionary standpoint, CaaD was of interest because it introduced a new activity to the tautomerase family – that of dehalogenation by hydration. Thus far, the characterized members of the tautomerase family including 5-carboxymethyl-2-hydroxymuconate isomerase (CHMI), macrophage migration inhibitory factor (MIF), 4-OT, and YdcE, a 4-OT homologue from *Escherichia coli* K-12, have only been known to catalyze tautomerization and/or isomerization reactions (5, 19, 20). CaaD, which shows sequence identity with 4-OT and YwhB, a homologue of unknown function from *Bacillus subtilis*, was initially placed in the tautomerase superfamily by virtue of its sequence similarity. Catalytic residues of interest include the β-Pro-1 and the α-Arg-11, which have been shown in other work to be crucial for catalysis (1). The N-terminal proline is also apparently conserved in the α-subunit of CaaD, although initial mutagenesis studies described elsewhere suggested it was not essential for catalysis (1). These observations raised questions about the role of the amino terminal prolines in the α- and β-subunits. The first question is whether β-Pro-1 functions as a general base catalyst, as it does in the characterized tautomerases, or as an acid catalyst? A second question is whether the conserved α-Pro-1 plays any role in the reaction. To answer these questions, we investigated the properties of α-P1A-CaaD and determined the pH rate profiles of both the wild-type and the α-P1A mutant enzymes.

In previous work, the α-P1A CaaD mutant was constructed but not rigorously characterized (1). A crude extract of mutant protein, generated by lysis and centrifugation, was used without further purification in a colorimetric assay that monitors halide production. Two measurements at a single substrate concentration (5 mM) of either **1** or **2** provided a specific activity, expressed in terms of units (the amount of enzyme that catalyzes the production of 1 μmol halide ion per minute).

The relative activities, compared to a similar procedure performed with wild-type enzyme, were 0.06-fold lower for **1** and 0.16-fold for **2**. On the basis of these observations, the authors suggested that α-Pro-1 was not necessary for catalysis. In view of the facts that the α-P1A mutant was not purified and kinetic parameters were not obtained, this mutant was re-examined.

The α-P1A mutant was made using site-directed mutagenesis. The α-P1A plasmid was co-expressed with the wild-type β-subunit plasmid, and the enzyme was purified to near homogeneity using the protocol developed for wild-type CaaD. DNA sequencing of the construct demonstrated that no other mutations had been introduced into the α-subunit plasmid. Mass spectrometry of the mutant protein indicated that the amino-terminal residues were not blocked by the initiating methionine. The yield of the mutant protein was slightly lower than that of the wild-type enzyme, but the purity was comparable as evaluated by SDS-PAGE.

UV assays have been developed to assess the activity of wild-type CaaD (13). One assay monitors the decrease in absorbance at 224 nm of the haloacrylate compounds **1** and **2** in phosphate buffer. Another substrate, 2-oxo-3-pentynoate (2-OP, **5**) was used to assess the ability of CaaD to catalyze a hydration reaction. The resulting product, acetopyruvate, shows a characteristic increase in absorbance at 294 nm. Hence, the three substrates were used to characterize the activity and properties of α-P1A-CaaD.

Accordingly, the steady-state parameters for the α-P1A-CaaD mutant were determined using trans-3-bromoacrylate (**1**), trans-3-chloroacrylate (**2**), and 2-OP (**5**) as substrates. The values for $k_{cat}$, $K_m$, and $k_{cat}/K_m$ that result from the UV assays are given in Table 5 together with those of the wild-type enzyme. The $k_{cat}$ and $k_{cat}/K_m$ values for the haloacrylate substrates **1** and **2** are slightly lower than observed for the wild-type, representing approximately 0.4 to 0.2-fold lower than the wild-type activity. The $K_m$ values for these substrates are approximately equal. These values suggest that although it may play a structural or catalytic role, α-Pro-1 is not an essential residue.

For the hydration of 2-OP (**5**), the values for $k_{cat}$ are identical. This suggests that α-Pro-1 is not crucial for the hydration reaction. The most striking difference observed is in the $K_m$ value for **5**, which is half that observed for the wild-type enzyme. The difference in $K_m$ values is also reflected in the values for $k_{cat}/K_m$. Replacement of proline with an alanine residue, which removes a larger ring system in favor of a smaller methyl group, may allow **5** to better fit into the active site as it is a larger molecule than either **1** or **2**. Thus, α-Pro-1 may play a role in maintaining the structural integrity of the entrance to the active site. In the absence of a structure for CaaD, a structural role for α-Pro-1 cannot be determined.

The kinetic pH rate profile for wild-type CaaD was determined in an attempt to elucidate the role of β-Pro-1. Two potential roles for β-Pro-1 have been proposed (1). In one mechanism, the proline could function as a general base in catalysis, deprotonating water for attack on the double bond. This hypothesis is consistent with the role of Pro-1 in other tautomerase superfamily members. In 4-OT, Pro-1 is able to act as a catalytic base because its p$K_a$ is lowered approximately 1000-fold to 6.4 (6, 7). Thus, at physiological pH (~7-7.3), 80-88% of the N-terminal prolines are deprotonated. If β-Pro-1 was a general base in the CaaD-catalyzed reaction, a similar drop in one of the p$K_a$s would be expected. For instance, an observed p$K_a$ of 7.5 would yield 24-39% deprotonation at physiological pH. In the second mechanism, β-Pro-1 functions as a general acid from which the proton at C-2 is obtained. In this event, a decrease in the p$K_a$ of β-Pro-1 would not be anticipated. However, the absence of a low p$K_a$ value would only be indirect evidence that β-Pro-1 is a general acid catalyst. Definite proof requires the assignment of the observed p$K_a$ value(s) to specific residues.

The pH rate profile for the wild-type enzyme was determined in 20 mM sodium phosphate buffer over the range of 6.5-10.26 at 30 °C. The initial pHs examined varied at 0.25 pH increments. However, the large amounts of enzyme required to detect activity altered the pH, particularly under fairly acidic or fairly basic conditions. Thus, the final pHs are not at regular intervals due to the changes

caused by the addition of enzyme. The pH dependence of log $k_{cat}/K_m$ (Figure 35) for the wild-type enzyme showed two limbs, each with a slope of unity. The slopes of one indicate that a single group is involved in base catalysis, and a single group is involved in acid catalysis. These groups may arise from either the free enzyme or the free substrate. The basic group, which corresponds to the ascending limb, has a p$K_a$ of 8.0 ± 0.2. To confirm that this value was not attributable to the carboxylate ion of (**1**), the p$K_a$ of the substrate was directly measured to be 4.6. Compound **1** has no ionizable groups near pH ~8, so the observed p$K_a$ does not correspond to that of the substrate. Similarly, the p$K_a$ of the descending limb indicates that an acidic group is necessary for activity. The p$K_a$ for this group is 9.1 ± 0.2.

These results are not consistent with β-Pro-1 functioning as a catalytic base. The p$K_a$ at 9.1 is comparable to the p$K_a$ (~ 9.4) of the model compound, proline amide. Proline amide is not a base at physiological pH since the prolyl nitrogen is nearly fully protonated. If β-Pro-1 were responsible for the observed p$K_a$ value at 9.1, it would be 98-99% protonated and could not act as a base under cellular conditions. If β-Pro-1 were responsible for the p$K_a$ at 8.0, approximately 83-91% of the proline would be protonated at physiological pH. While this represents a significant increase in the amount of available base, it is unlikely that β-Pro-1 would be an efficient general base catalyst.

In order to determine whether one of the p$K_a$ values was due to α-Pro-1, we examined the pH dependence of the α-P1A-CaaD reaction. The final pH values vary, again due to the amount of enzyme added at each pH. The results, however, are strikingly similar to those of the wild-type. The observed p$K_a$ values for $k_{cat}/K_m$ (Figure 36) are 7.9 ± 0.2 and 9.0 ± 0.2 on the ascending and descending limbs, which both show a slope of one. Thus, a single basic group and a single acidic group from the free enzyme or the free substrate are involved in catalysis. Again, the substrate can be ruled out because its p$K_a$ is 4.6 and is not within the range observed here. Proline-1 of the α-subunit is also not responsible for either value. If α-Pro-1 were responsible for one of the values, a change would have been observed. However, the

corresponding wild-type p$K_a$ values were 8.0 ± 0.2 and 9.1 ± 0.2, which are equivalent to those observed for the mutant.

Similarities are also evident in the pH dependences of log $k_{cat}$ for the wild-type and mutant enzymes. The p$K_a$ values for $k_{cat}$ are 6.8 ± 0.2 and 9.6 ± 0.2 (Figure 35) for the wild-type and 6.4 ± 0.2 and 9.7 ± 0.2 for α-P1A-CaaD (Figure 36). In each plot, the slope of the ascending and descending limbs is equal to one. Thus, a single basic residue and a single acidic residue are important for catalysis in the enzyme-substrate complex. The plot of log $k_{cat}$ is indicative of the behavior of the enzyme under saturating substrate conditions. Thus, any contribution by the substrate to the reaction can be ruled out. The equivalence of the p$K_a$ values obtained reinforces the earlier results that show that α-Pro-1 does not contribute to the reaction as either the acid or base catalyst.

From these analyses, it is likely that β-Pro-1 does not function as a base catalyst. However, the p$K_a$ values observed in the wild-type or α-P1A-CaaD pH rate profiles cannot be assigned to specific residues. A pH rate analysis of the β-P1A-CaaD mutant would have provided further insight into the p$K_a$ value for that particular residue. If the β-Pro-1 was responsible for either the p$K_a$ value at 8.0 or 9.1, the corresponding value would change upon substitution by an alanine. However, kinetic analysis of the β-P1A CaaD mutant was not possible due to the inactivity of the enzyme.

Thus, assignment of kinetically determined p$K_a$ values is performed by direct pH titration using NMR spectroscopy. Such an analysis requires uniformly $^{15}$N-labeled protein. Previous studies in the Whitman laboratory have used a complex procedure involving a MOPS-buffered medium for $^{15}$N-labeling using $^{15}NH_4Cl$ as a sole source of nitrogen that required several temperature and media changes over the course of two days (6). Despite the optimization of this process, the yield of isotopically labeled protein is typically less than that of non-isotopically labeled protein. One reason for this effect is the slower bacterial growth under minimal media conditions. LB medium, used for non-isotopically labeled protein, is a rich

medium that provides numerous carbon, nitrogen and amino acid resources. Minimal media provides carbon and nitrogen only in the forms of glucose and salts, eliminating amino acid, vitamin, and nucleotide sources. These conditions force the bacteria to synthesize not only proteins, but all of the necessary cofactors and nucleic acids required for growth using the minimal carbon and nitrogen sources.

In an effort to simplify the procedure, growth was performed in M9 minimal media that was enhanced with several supplements as described in the methods section. The new protocol requires only an initial overnight growth in LB at 37 °C, followed by direct inoculation into minimal media. The minimal media phase requires only 24 hr of incubation and induction at 30 °C. Comparable procedures in the literature (18) indicated that high yields of ~98% labeled protein could be obtained in this fashion. Initally, the growth of 4-OT in M9 minimal medium was performed as a control. The yield and purity of the resulting protein were excellent (~40 mg/L). Unfortunately, the same procedure used for CaaD resulted in low yields of $^{15}$N-isotopically labeled protein. This may be due to the additional burden of dual plasmid retention and expression, since cells limit the number of different plasmids they will maintain. Addition of nucleotides, vitamins, and trace mineral supplements allowed the bacteria to express labeled CaaD. Still, the overall yields diminished greatly, dropping to ~10% of typical non-isotopically labeled yields.

Despite these difficulties, sufficient quantities of the $^{15}$N-labeled wild-type and α-P1A enzymes were purified for initial NMR titration experiments. Although other work has shown that the $^{15}$N-chemical shift of the prolyl nitrogen in 4-OT is well resolved (7), titration of wild-type CaaD is complicated by the fact that both the α- and β-subunits have an amino-terminal proline. Hence, the $^{15}$N-labeled αP1A mutant was also examined. For the wild-type enzyme, NMR titration yielded two p$K_a$ values at 9.2 ± 0.1 and 11.1 ± 0.1. In the $^{15}$N NMR spectrum of the mutant enzyme, the α-Pro-1 is absent. NMR titration yielded a single p$K_a$ at 9.9 ± 0.2. The similarity of the p$K_a$ to that observed for the acidic group (9.1 ± 0.2 in wild-type CaaD and 9.0 ± 0.2 in α-P1A-CaaD) implicates β–Pro-1 as a general acid catalyst in the mechanism,

responsible for protonation at C-2 of **1**. An unknown basic residue, possibly having a p$K_a$ of 8.0 ± 0.2, is likely involved in the activation of water.

Although the role of β-Pro-1 in CaaD is reversed from that of the N-terminal prolines in other tautomerase superfamily members, it is not surprising in view of the evolutionary context and the nature of the reaction catalyzed. Most dehalogenases have likely evolved over the past few decades, which is quite recent on the evolutionary timescale. Thus, it is reasonable to assume that CaaD evolved from a progenitor that is closely related to 4-OT. Perhaps the active site of the progenitor was, like 4-OT, very hydrophobic. Phe-50 in 4-OT is crucial for maintaining the p$K_a$ of Pro-1 through hydrophobic interactions (21). The active site of CaaD, however, is likely to be hydrophilic because it catalyzes a hydration reaction. Mutations in the active site, affording greater hydrophilicity, resulted in an upward shift of the p$K_a$ of the β-Pro-1, forcing the enzyme to utilize the residue as an acid. CaaD represents a new activity as well as a new structure within the 4-OT family, and provides evidence for the catalytic and structural versatility of the β–α–β structural fold. The results of this study show that there is also versatility in the reactivity of Pro-1 within the 4-OT family: it can be utilized as either a general base or a general acid catalyst. Further elucidation of the principles used by Nature to manipulate the reactivity of key elements in the simple β–α–β motif may assist in the creation of new biocatalysts using this basic scaffolding.

**ACKNOWLEDGMENTS**

Electrospray ionization (ESI) mass spectrometry was performed by the Analytical Instrumentation Facility Core (College of Pharmacy, The University of Texas at Austin) supported by Center grant ES 07784. We thank Dr. Maria D. Person (Division of Pharmacology and Toxicology, The University of Texas at Austin) for obtaining mass spectra and Dr. William H. Johnson, Jr. (Division of Medicinal Chemistry, The University of Texas at Austin) for synthesizing needed compounds and interpreting the NMR spectra. We also thank Dr. Hugo F. Azurmendi, Dr. Michael A. Massiah, and Dr. Albert S. Mildvan (Department of Biological Chemistry, Johns Hopkins University) for providing NMR data prior to publication.

**REFERENCES**

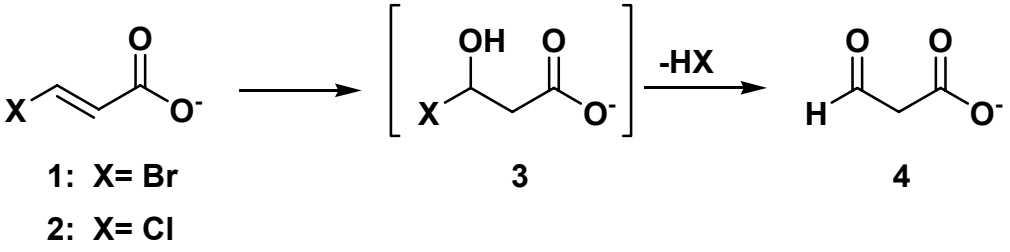

1. Poelarends, G. J., Saunier, R., and Janssen, D. B. *J. Bacteriol.,* 2001, **183**, 4269-4277.
2. Poelarends, G. J., Kulakov, L. A., Larkin, M. J., van Hylckama Vlieg, J. E. T., and Janssen, D. B. *J. Bacteriol.,* 2000, **182**, 2191-2199.
3. Poelarends, G. J., Wilkens, M., Larkin, M. J., van Elsas, J. D., Janssen, D. B. *Appl. Environ. Microbiol.* 1998, **64**, 2931-2936.
4. Whitman, C. P. *Arch. Biochem. Biophys.* 2000, **402**, 1-13.
5. Almrud, J. J., *et al. Biochemistry* 2002, **41**, 12010-12024.
6. Stivers, J. T., *et al*. *Biochemistry*, 1996, **35**, 803-813.
7. Stivers, J. T., Abeygunawardana, C., Mildvan, A. S., Hajipour, G., and Whitman, C. P. *Biochemistry,* 1996, **35**, 814-823.
8. Czerwinski, R. M., Harris, T. K., Massiah, M. A., Mildvan, A. S., and Whitman, C. P. *Biochemistry,* 2001*,* **40**, 1984-1995.
9. Sambrook, J., Fritsch, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual, 1989.
10. Neidhardt, F. C., Block, P. L., and Smith, D. F. *J. Bact.*, 1974, **119**, 736-747.
11. Andersson, K. *Chem. Scripta*, 1972, **2**, 117-120.
12. Johnson, W. H., Jr., Czerwinski, R. M., Fitzgerald, M. C., and Whitman, C. P. *Biochemistry,* 1997, **36**, 15724-15732.
13. Wang, S. C., Person, M. D., Johnson, W. H., Jr., and Whitman, C. P. *Biochemistry*, 2003, **42**, 8762-8773.
14. Laemmli, U. K. *Nature*, 1970, **227**, 680-685.
15. Waddell, W. J. *J. Lab. Clin. Med.,* 1956, **48**, 311-314.
16. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. *Gene,* 1989, **77**, 51-59.
17. Marcotte, P., and Walsh, C. T. *Biochemistry,* 1978, **17**, 5613-5619.
18. Mayer, K. L., and Stone, M. J. *Biochemistry*, 2000, **39**, 8382-8395.

19. Whitman, C. P., *et al. JACS,* 1992, **114**, 10104-10110.

20. Stamps, S. L., Fitzgerald, M. C., and Whitman, C. P. *Biochemistry*, 1998, **37**, 10195-10202.

21. Czerwinski, R. M., Harris, T. K., Massiah, M. A., Mildvan, A. S., and Whitman, C. P., *Biochemistry*, 2001, **40**, 1984-1995.

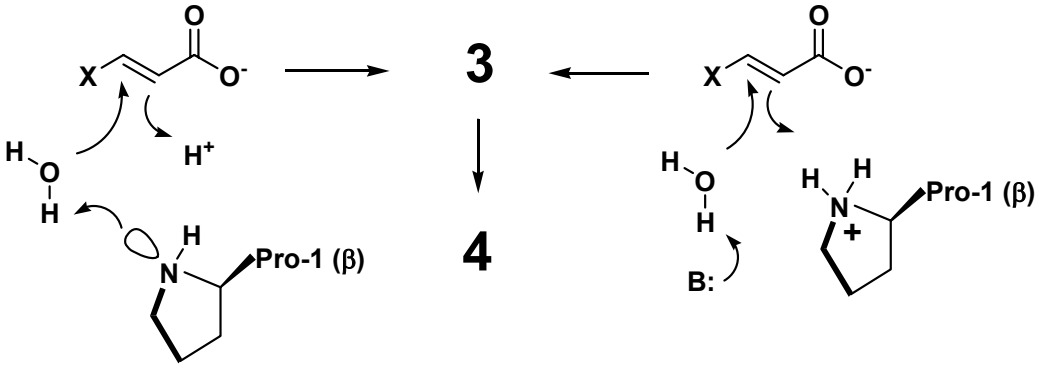
**Scheme 19**



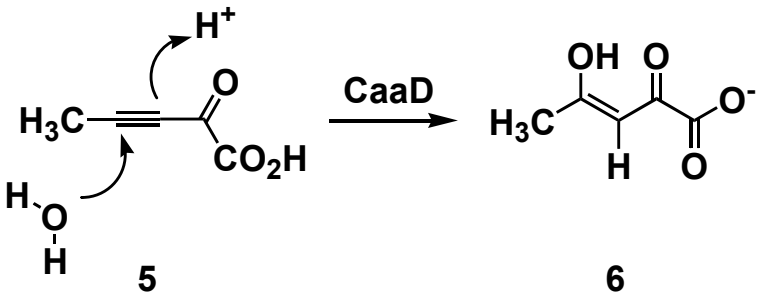

**1: X= Br** **3** **4**

**2: X= Cl**

**Scheme 20**



**Scheme 21**




5 6

**Figure 35**. pH dependence of the kinetic parameters for the CaaD-catalyzed dehalogenation of **1**. The pH dependence of log ($k_{cat}/K_m$) (open circles) and log $k_{cat}$ (filled circles) are shown. The curves were computed from a non-linear least-squares fit as described elsewhere (6). The p$K_a$ values are 8.0 ± 0.2 and 9.1 ± 0.2 for log ($k_{cat}/K_m$) and 6.8 ± 0.2 and 9.6 ± 0.2 for log $k_{cat}$.

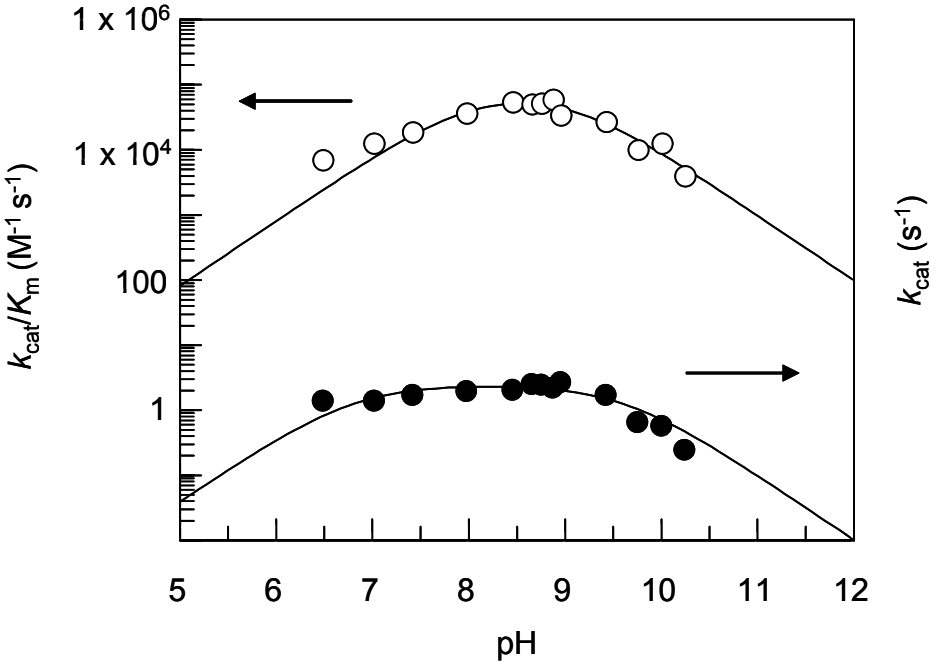


**Figure 36.** pH dependence of the kinetic parameters for the α-P1A-CaaD-catalyzed dehalogenation of **1**. The pH dependence of log ($k_{cat}/K_m$) (open circles) and log $k_{cat}$ (filled circles) are shown. The curves were computed from a non-linear least-squares fit as described elsewhere (6). The p$K_a$ values are 7.9 ± 0.2 and 9.0 ± 0.2 for log ($k_{cat}/K_m$) and 6.4 ± 0.2 and 9.7 ± 0.2 for log $k_{cat}$.

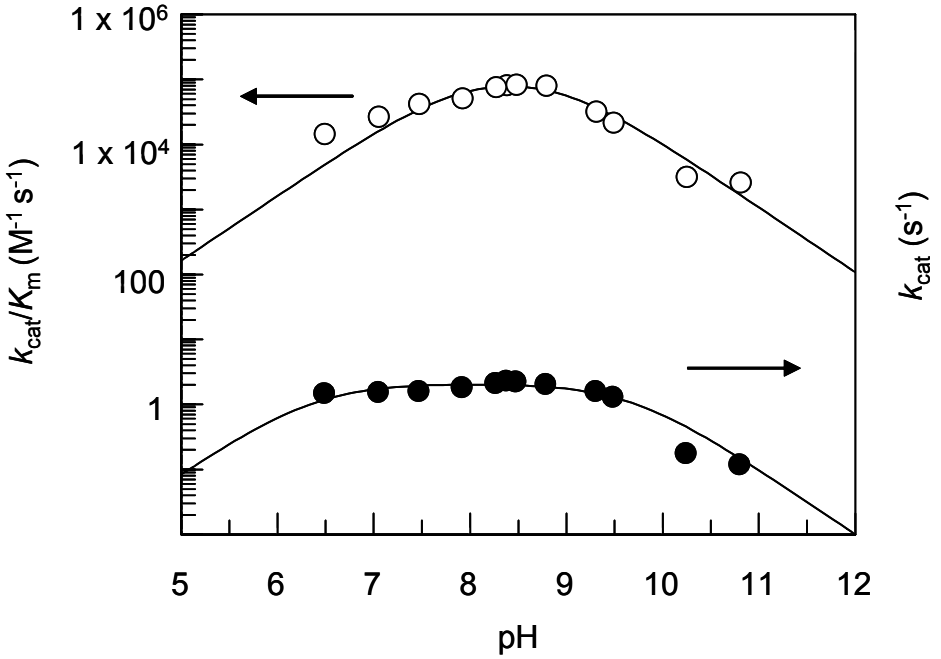


**Table 5. Kinetic Parameters for CaaD and α-P1A-CaaD[a]**

| enzyme | substrate | $k_{cat}$ ($s^{-1}$) | $K_m$ (μM) | $k_{cat}/K_m$ ($M^{-1}$ $s^{-1}$) |
|---|---|---|---|---|
| CaaD | **1** | 5.1 ± 0.1 | 37 ± 2 | $1.4 \times 10^5$ |
| | **2** | 3.8 ± 0.1 | 31 ± 2 | $1.2 \times 10^5$ |
| | **5** | 0.7 ± 0.01 | 110 ± 4 | $6.4 \times 10^3$ |
| α-P1A-CaaD | **1** | 3.9 ± 0.1 | 43 ± 3 | $9.1 \times 10^4$ |
| | **2** | 2.5 ± 0.1 | 33 ± 2 | $7.4 \times 10^4$ |
| | **5** | 0.7 ± 0.01 | 49 ± 2 | $1.4 \times 10^4$ |

[a]The steady-state kinetic parameters were determined in 20 mM potassium phosphate buffer (pH 9.0) at 23 °C. Errors are standard deviations. The values for CaaD have previously been reported (13).

## KINETIC CHARACTERIZATION OF YWHB

### INTRODUCTION

4-Oxalocrotonate tautomerase (4-OT) catalyzes the conversion of 2-oxo-4-hexenedioate (**1**, Scheme 22) to 2-oxo-3-hexenedioate (**3**) via the intermediate, 2-hydroxy-2,4-hexadienedioate, known commonly as 2-hydroxymuconate (**2**). The enzyme is a hexamer, made up of monomers consisting of 62 amino acids, and is part of the catechol meta-fission pathway, which is a degradative pathway found on the TOL plasmid. Bacteria such as *Pseudomonas putida mt-2* that harbor the TOL plasmid can utilize various aromatic hydrocarbons as their sole sources of carbon and energy.

The mechanism and structure of 4-OT have been extensively studied over the past decade. On the basis of affinity labeling, NMR, kinetic, mutagenesis, crystallographic, and chemical synthesis studies, the amino-terminal proline of 4-OT has been implicated as the catalytic base responsible for the observed 1,3-suprafacial allylic rearrangement (1-8). In addition, Arg-11 and Arg-39 play critical roles (9). Arg-39 binds the C-1 carboxylate group of **1** and provides one of the two hydrogen bonds to the carbonyl oxygen thereby facilitating catalysis, while Arg-11 binds the C-6 carboxylate group and functions as an electron sink. An ordered water molecule, held in place by Ser-37, provides a second hydrogen bond to the carbonyl oxygen (5). Finally, Phe-50 is largely responsible for maintaining the hydrophobic nature of the active site, which lowers the p$K_a$ of Pro-1 to ~6.4, enabling it to function as a general base catalyst (10).

4-OT is also the best-characterized member of the 4-OT family, one of three major families comprising the tautomerase superfamily (11, 12). The members of this superfamily are structurally homologous proteins sharing a characteristic β–α–β structural motif as well as a conserved amino-terminal proline (11). We recently identified 38 4-OT homologues in bacteria ranging from common organisms such as

*Escherichia coli* and *Bacillus subtilis* to more unusual organisms such as the thermophiles *Archaeoglobus fulgidus* and *Methanobacterium thermoautotrophicum* (13). In contrast to 4-OT, these homologues are chromosomally encoded and most are found in organisms not known to degrade aromatic hydrocarbons.

The 4-OT homologue in *B. subtilis*, designated YwhB, is a closely related homologue, sharing ~36% sequence similarity. The 61-amino acid YwhB sequence is a hexamer due to the presence of an AVAG sequence (residues 51-54), which generates a β-hairpin that is critical for hexamer formation (13). The YwhB sequence retains the catalytic amino-terminal proline as well as Arg-11, and replaces Phe-50 with a tyrosine. A lysine residue replaces Ser-37. The only "missing" catalytic residue in YwhB is Arg-39, which is replaced with a valine.

YwhB also shares ~35% sequence similarity with the α-subunit of *trans*-3-chloroacrylic acid dehalogenase (CaaD) from *Pseudomonas pavonaceae strain 170* (14). CaaD, a heterohexamer consisting of three α-subunits and three β-subunits, catalyzes the addition of water to the *trans* isomers of 3-bromo- and 3-chloroacrylate to form malonate semialdehyde (14). Pro-1 of the β-subunit and Arg-11 of the α-subunit have been identified as critical residues in the mechanism of CaaD (14).

These sequence similarities suggested that YwhB might be the precursor to both 4-OT and the α-subunit of CaaD and prompted us to clone, express, and conduct a kinetic and stereochemical analysis of YwhB. The enzyme functions as an effective tautomerase, converting **2** to **1** (Scheme 22), 2-hydroxy-2,4-pentadienoate (**5**) to its respective β,γ-unsaturated isomer, 2-oxo-4-pentenoate (**4**, Scheme 22), the enol isomers of phenylpyruvate (**7**) and (*p*-hydroxyphenyl)pyruvate (**9**), to their respective keto isomers (**8** and **10**, Scheme 23), and 2-hydroxy-2,4-heptadiene-1,7-dioate (**11**) to 2-oxo-4-heptene-1,7-dioate (**12**). Despite its close resemblance to 4-OT, YwhB processes **2** to **3** only very slowly. While these results demonstrate that YwhB has a tautomerase activity, the role this activity plays in the organism is not clear. A more interesting question may be how this activity was transformed into 4-OT or the α-subunit of CaaD, if YwhB is truly their progenitor.

**MATERIALS AND METHODS.**

*Materials*. All reagents, buffers, and solvents were obtained from Sigma Aldrich Chemical Co. (St. Louis, MO) or Fisher Scientific Inc. (Pittsburgh, PA). Tryptone, yeast extract, and agar were obtained from Becton, Dickerson, and Company (Franklin Lakes, NJ). The Ultrafree DA centrifugal filter units, the YM-3 ultrafiltration membranes, and the Amicon stirred concentrators were obtained from Millipore Corp. (Billerica, MA). Thin-walled PCR tubes were obtained from Ambion, Inc. (Austin, TX). The syntheses of 2-hydroxymuconate (**2**), 2-hydroxy-2,4-pentadienaote (**5**), and 2-hydroxy-2,4-heptadiene-1,7-dioate (**11**) are described elsewhere (15, 16). The expression vector pET-24a(+) was obtained from Novagen, Inc. (Madison, WI). Restriction enzymes, T4 DNA ligase, PCR reagents, agarose, the Wizard PCR Preps DNA purification kit, and the Wizard Plus Minipreps DNA Purification System were obtained from Promega Corp. (Madison, WI). Oligonucleotides for DNA sequencing, amplification, or mutant construction were synthesized by Oligos Etc. Inc (Wilsonville, OR) or Genosys (The Woodlands, TX).

*Strains.* *Escherichia coli* strains JM109 (Promega), DH5α (Invitrogen Corp. Carlsbad, CA), and BL21-Gold(DE3)pLysS (Stratagene, La Jolla, CA) were used for transformation of ligated plasmids. *E. coli* strains BL21(DE3)pLysS (Novagen) and BL21-Gold(DE3)pLysS were used for expression of the recombinant proteins. Cells for general cloning and expression were grown in LB media supplemented with kanamycin (50-100 μg/mL) or kanamycin and chloramphenicol (25 μg/mL). The composition of LB medium is described elsewhere (17).

*General Methods.* Techniques for restriction enzyme digestions, ligation, transformation, and other standard molecular biology manipulations were based on methods described elsewhere (17). DNA sequencing was done at the University of Texas (Austin) Sequencing Facility. Kinetic data were obtained on a Hewlett Packard 8452A or an Agilent 8453 Diode Array spectrophotometer. The cuvettes

were mixed by a stir/add cuvette mixer (Bel-Art Products, Pequannock, NJ). The kinetic data were fitted by nonlinear regression data analysis using the Grafit program (Erithacus Software Ltd., Staines, U. K.) obtained from Sigma Chemical Co. High performance liquid chromatography (HPLC) was performed on a Waters (Milford, MA) 501/510 system or a Beckman System Gold HPLC (Fullerton, CA) using a TSKgel Phenyl-5PW hydrophobic column (Tosoh Bioscience, Montgomeryville, PA). Protein was analyzed by Tris glycine sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions on 17.5% gels on a vertical gel electrophoresis apparatus obtained from Bio-Rad (Hercules, CA) (18). Protein concentrations were determined using the method of Waddell (19). NMR spectra were obtained on a Bruker AM-250 spectrometer or a Bruker AM-500 spectrometer as indicated. Chemical shifts were referenced as noted below.

*Construction of the Expression Vector for YwhB from B. subtilis.* The expression vector for YwhB was designed in previous work by Dr. Robert M. Czerwinski (manuscript in preparation). The plasmid was designated pET-24a-ywhB. Subsequently, the plasmid containing the YwhB gene was transformed as described elsewhere into *E. coli* strains BL21(DE3)pLysS or BL21-Gold(DE3)pLysS for protein expression according to the manufacturer's directions.

*Overexpression, Purification, and Characterization of the YwhB Product***.** The product of the YwhB gene was overexpressed using a previously described protocol (10). Typically, 4.5 L of culture grown under these conditions yields ~14 g of cells. The enzyme was purified to near homogeneity (>99% as assessed by the presence of a single band in SDS-PAGE) by a procedure described elsewhere using both the Tosoh Phenyl 5-PW hydrophobic column and the Sephadex G-75 gel filtration column (2.5 × 100 cm) using 20 mM $NaH_2PO_4$ buffer (pH 7.3). Typically, the yield of purified protein per liter of culture was ~50 mg. The native molecular mass of the protein was estimated by size exclusion chromatography on a Superose 12 column. The enzyme was chromatographed in 100 μL portions (~6.2 mg/mL) on the column equilibrated with 20 mM $NaH_2PO_4$ buffer, pH 7.3, at a flow rate of 0.2

mL/min. The protein was monitored at 280 nm and typically elutes at ~74 min. The molecular mass ($MH^+$) for the YwhB product was 7015.0 Da (calc. 7014.0 Da) as determined by ESI-MS.

*Construction of K37A and K37A V39R YwhB.* The expression vectors for two YwhB mutants, K37A and K37A V39R, were constructed using overlap extension PCR by Dr. Robert M. Czerwinski (unpublished data). These plasmids were designated pET-24a ywhB K37A and pET-24a ywhB K37A V39R, respectively, and transformed into *E. coli* strain BL21(DE3)pLysS for protein expression according to the manufacturer's instructions.

*Construction of P1A, R11A, and V39R YwhB*. The P1A, R11A, and V39R mutants of YwhB were generated using the gene for YwhB in the plasmid pET-24a-ywhB as the template. The gene is flanked by an *Nde*I restriction site and a *Not*I restriction site (one of the multiple restriction sites). The mutations were made using the overlap extension polymerase chain reaction as described (20). The external PCR primers were oligonucleotides 5'-GCG GAT AAC AAT TCC CCT CT-3' (designated primer A) and 5'-CTC AGC TTC CTT TCG GGC TT-3' (designated primer D), which were described in Chapter 3. For the P1A mutant, the internal PCR primers were 5'-GAC AGT TAC GTA TGC CAT ATG TATA- 3' (primer B) and 5'-TAT ACA TAT GGC ATA CGT AAC TGT C-3' (primer C). For the R11A mutant, the internal PCR primers were 5'-TTG TTC GTC TGT AGC GCC TTC GAG CAT-3' (primer B) and 5'- ATG CTC GAA GGC GCT ACA GAC GAA CAA-3' (primer C). For the V39R mutant, the internal PCR primers were oligonucleotides 5'-TTC TAT AAA GAC ACG AAT TTT TTC TTC - 3' (primer B) and 5' – GAA GAA AAA ATT CGT GTC TTT ATA GAA - 3' (primer C). In each set of primers, the mutation is underlined and the remaining bases correspond to the coding sequence (primer C) or the complementary sequence (primer B).

PCR reactions were carried out as described for amplification of the wild type gene with the following modifications. In two separate PCRs, the AB and CD fragments were constructed in reaction mixtures containing 10× buffer (5 μL), the

dNTPs (1 μL of a 10 mM stock solution), primers (2.5 μL of 20 μM stock solutions), $MgCl_2$ (3 μL of a 25 mM solution), pET-24a-ywhB (1 μL of a 1:500 dilution of the plasmid), *Taq* DNA polymerase (0.5 μL of a 5 units/μL solution) and a sufficient volume of water to make a final volume of 50 μL. The reaction protocol consisted of 25 cycles of PCR (denaturation at 94 °C for 1 min, annealing at 55 °C for 75 s, and elongation at 72 °C for 75 s), preceded by an initial denaturation step at 94 °C for 5 min and followed by a final elongation step at 72 °C for 5 min. The PCR products were analyzed by electrophoresis on a 1% agarose gel and purified using either the Wizard PCR Preps purification system. The AD fragments were constructed similarly except that the AB and CD fragments (2.5 μL of each) were used as template.

The AD fragments were electrophoresed, excised, and purified using the Wizard PCR Preps or the Ultrafree-DA centrifugal filter units. In separate reactions, the PCR products and the pET-24a(+) vector were treated with *Nde*I and *Not*I for 5-7 hrs at 37 °C. Ligation using T4 DNA ligase was performed at room temperature overnight. Subsequently, the ligated DNA was purified by ethanol precipitation and used to transform DH5α cells or BL21-Gold(DE3) pLysS by electroporation. Colonies (grown on either LB/Kn or LB/Kn/Cm agar plates) were randomly selected for plasmid purification and DNA sequencing. The plasmids designated pET-24a-ywhB P1A-4, pET-24a-ywhB R11A-A, and pET-24a-ywhB V39R-4 had the correct sequence and were used in subsequent experiments.

*Overexpression and Purification of the YwhB Mutants.* All of the mutants were overexpressed and purified in the same manner as the wild-type. Elution times for all of the mutants from both columns were approximately the same as those observed for the wild type. The typical yields of cells from 3 L of culture were 10-14 g for each mutant. The yields of each mutant were comparable to that of the wild type (20-40 mg/L culture).

*Kinetic Characterization of YwhB and the Mutants*. Kinetic runs were performed in 20 mM $NaH_2PO_4$ buffer, pH 7.3. Stock solutions (50 mM) of **2**, **10**, and

**11** were initially made up in ethanol, and diluted (with ethanol) to generate 5 mM and/or 10 mM solutions. Solutions of **5** were made by diluting a stock solution (152.6 mM) stored in ethanol at -20° C, to give 3.05 mM and 30.5 mM solutions. The tautomerization activity of YwhB was measured by monitoring the ketonization of the following substrates at the indicated $\lambda_{max}$ and $\varepsilon$ values, as follows: **2** (294 nm, $\varepsilon$ = 19,400 $M^{-1}$ $cm^{-1}$), **5** (266 nm, $\varepsilon$ = 12,100 $M^{-1}$ $cm^{-1}$), **7** (288 nm, $\varepsilon$ = 18,000 $M^{-1}$ $cm^{-1}$), **9** (292 nm, $\varepsilon$ = 20,800 $M^{-1}$ $cm^{-1}$), **14** (276 nm, $\varepsilon$ = 12,300 $M^{-1}$ $cm^{-1}$), and described elsewhere (1, 15, 16). The formation of **3** from **2** was monitored at 236 nm ($\varepsilon$ = 6580 $M^{-1}$ $cm^{-1}$) (1). The ranges of substrates were as follows for ketonization: **2** (20-150 μM), **5** (3-150 μM), **10, 11,** and **14** (10-150 μM). For isomerization, the range of **2** was 30-600 μM. For a typical run, a sufficient quantity of YwhB as required for each particular activity was equilibrated in buffer (typically 4 μL of a 25.6 mg/mL solution in 40 mL for **2**, 40 μL of a 25.6 mg/mL solution in 40 mL for **14**) for at least an hour prior to use. Aliquots (1 mL) were withdrawn and assayed for activity. Reactions were initiated by the addition of substrate.

## RESULTS

*Production, Expression, and Characterization of YwhB and the Mutants.* In work described elsewhere, YwhB was cloned using *Bacillus subtilis* strain 168 genomic DNA as template (unpublished data). The construction of the expression system and the purification protocol followed those used for 4-OT (10). Accordingly, YwhB can be purified to homogeneity (~99%), and elutes under the same conditions as 4-OT. The yield of protein is high (~50 mg/L) of culture. Mass spectral analysis of the purified wild-type protein indicates that it has the expected mass, and that the N-terminal proline is not blocked by the initiating methionine. The native molecular mass as estimated by gel filtration indicates that YwhB exists as a hexamer in solution. This has been confirmed by crystallographic studies (12). Expression and purification of all of the mutants was carried out similarly. All of the mutant proteins remained soluble and were readily purified with yields ranging from 50-100% of wild-type.

*Kinetic Properties of the YwhB Protein.* The kinetic properties of YwhB were investigated using as potential substrates a monocarboxylated dienol (**5**), two enols (**7** and **9**), and two dicarboxylated dienols (**2** and **11**). For the four dienol substrates, the rates of formation of the corresponding β,γ-unsaturated ketones were monitored. For **5** and **11**, the rates of formation of the corresponding α,β-isomers were not measurable. For **2**, however, it was possible to determine a rate of formation for **3**. The kinetic data are summarized in Table 6.

Several trends emerge from the data. First, YwhB clearly catalyzes an 1,3-keto-enol tautomerization reaction using all of the substrates although with different efficiencies. Second, the highest catalytic efficiencies are observed with the monocarboxylated dienol **5**, the enol **7** (a 2.4-fold decrease in $k_{cat}/K_m$), and the dicarboxylated **2** (a 7.7-fold decrease in $k_{cat}/K_m$). Replacing the vinyl group of **5** with a phenyl group (i.e., **7**) has little effect on $k_{cat}$ and $K_m$. The addition of a second carboxylate group to **5** (i.e., **2**), results in a 17.8-fold decrease in $k_{cat}$, as well as a 2.3-

fold decrease in $K_m$. The decrease in catalytic efficiency observed for 2 is primarily a reflection of the decrease in $k_{cat}$. Third, the addition of the hydroxyl group to **7** results in a 30-fold decrease in $k_{cat}$ as well as a 2-fold decrease in $K_m$. The overall effect is a 38-fold decrease in the value of $k_{cat}/K_m$ (compared with **5**). Finally, the addition of a carboxymethyl group to **5** (i.e., **11**) results in a 59-fold decrease in $k_{cat}$ , a 1.9-fold increase in $K_m$, and an overall 111-fold decrease in $k_{cat}/K_m$.

The data in Table 6 also show that YwhB will catalyze a 1,5-keto enol tautomerization using **2**, although not as efficiently as 4-OT. 4-OT catalyzes the conversion of **2** to **3** with a $k_{cat}$ = 3500 $s^{-1}$ and a $K_m$ of 180 μM, resulting in an overall $k_{cat}/K_m$ of 1.9 x $10^7$ $M^{-1}$ $s^{-1}$. YwhB shows a 135-fold decrease in $k_{cat}$, a 5.2 increase in $K_m$, and a 680-fold decrease in $k_{cat}/K_m$. Interestingly, the kinetic parameters obtained for YwhB are on the order of those obtained for the R39A mutant of 4-OT ($k_{cat}$ = 28 s-1, $K_m$ of 290 μM, and $k_{cat}/K_m$ of 9.7 x $10^4$ $M^{-1}$ $s^{-1}$) (8). Moreover, replacement of R39 in 4-OT with a bulkier glutamine group results in a further increase in $K_m$ (to 470 μM). Thus, the presence of Val-39 is likely to increase the $K_m$ for **2**.

*Kinetic Properties of the P1A, R11A, and V39R Mutants of YwhB*. The purpose of the mutagenesis studies was two-fold. First, mutants of YwhB were constructed to determine whether the conserved catalytic residues (e.g., Pro-1 and Arg-11) were critical to the 1,3- and 1,5-keto-enol tautomerization reaction catalyzed by YwhB. Second, it was of interest to determine how many mutations would be necessary to convert YwhB into 4-OT. Accordingly, the P1A, R11A, and V39R mutants were analyzed using **2**, **7**, and **9** (Tables 7-10).

The data indicate that both the P1A and R11A mutants of YwhB are critical in catalysis. Both mutants are completely unable to process **2**. Neither a tautomerization or a ketonization ability using **2** is observed. The corresponding mutants of 4-OT show decreased activity towards **2** (32-fold and 760-fold decreases in $k_{cat}/K_m$, respectively), but the activity remains observable (8, 22). In the case of **7**, which is the best substrate of the three investigated with the mutants, the $K_m$ values improved with the mutation (1.5-fold and 3.7-fold for P1A and R11A, respectively).

The $k_{cat}$ values decreased significantly, by 100-fold for P1A and 2300-fold for R11A. Thus the differences in $k_{cat}$ values are largely reflected in the decreased values for $k_{cat}/K_m$ observed (71-fold and 610-fold), which more closely approximate the differences in $k_{cat}/K_m$ observed in the P1A and R11A 4-OT mutants. Finally, for **9**, neither mutant can catalyze the keto-enol tautomerization.

Replacement of Val-39 with an arginine residue results in a decrease in the ability of the mutant enzyme to process the phenyl compound **7**. A 2.6-fold decrease is observed in $k_{cat}$. However, only a 1.6-fold decrease is observed in $k_{cat}/K_m$ as a result of the improvement in $K_m$ (1.6-fold). The kinetic parameters for **9** as substrate largely remain similar. The values for $k_{cat}$ are equal, while a decrease in $k_{cat}/K_m$ (1.4-fold) can be attributed to a similar increase in $K_m$. An increase in the ability of the V39R mutant to process **2** to either **1** or **3** is also observed. A 21-fold improvement is observed in $k_{cat}$ for the 1,3-keto-enol tautomerization of **2** to **1**. The $K_m$ also increases 2.1-fold, thus leading to an overall 10-fold rate enhancement. The $k_{cat}$ and $K_m$ values are also enhanced in the 1,5-keto-enol tautomerization by 3-fold and 1.7-fold, respectively, thus leading to an overall 5-fold rate enhancement.

*Kinetic Properties of the K37A and K37A V39R Mutants of YwhB.* The K37A mutant was constructed in order to determine the importance of the lysine residue to catalysis. Since the V39R mutant showed an improved ability to function as a 4-OT, the K37A V39R double mutant was constructed in order to determine whether it would show an increased enhancement compared to the V39R mutation alone. If this were the case, 4-OT may have evolved from YwhB as a result of two primary mutations. Accordingly, the K37A and K37A V39R mutants were analyzed using **2**, **7**, and **9** (Tables 2-5).

The results indicate that lysine-37 carries out an important catalytic function. The K37A mutation results in changes for all of the kinetic parameters. Most notably, the ability of the mutant to carry out the 1,5-keto-enol tautomerization of **2** is abolished. K37A retains a very limited ability to catalyze the 1,3-keto-enol tautomerization, resulting in an overall 109-fold decrease in rate enhancement. This

change is less than that observed for the R39A and R39Q mutants of 4-OT (200-fold and 1000-fold, respectively). However, it appears that Lys-37 may fill a role in YwhB that is similar to that of Arg-39. Using **7** and **9** as substrates, changes remain apparent but are less striking. A 10-fold decrease in the tautomerization of **7** is observed in both $k_{cat}$ and $k_{cat}/K_m$, although $K_m$ improves slightly (1.4-fold). A similar 7.7-fold decrease in the rate enhancement for the tautomerization of **9** is observed. In the case of **11**, both $k_{cat}$ and $K_m$ suffer (5.9-fold and 1.2-fold, respectively).

The K37A V39R mutant did not achieve an improvement of the 4-OT functionality of YwhB. The results of the double mutation indicate drops in all of the activities investigated. The only improvements observed were a 2-fold increase in $k_{cat}$ for the 1,3-keto-enol tautomerization of **2**, and a 1.7-fold decrease of $K_m$ with **9** as substrate. In all other cases, the kinetic parameters were degenerated to varying degrees (50- to 5-fold in $k_{cat}$, 2.9- to 3.5-fold in $K_m$, and 1.7- to 27-fold in $k_{cat}/K_m$). The ability of the double mutant to tautomerize **9** was abolished.

## DISCUSSION

When 4-OT was first cloned and overexpressed in 1992, it was thought to be unique because no homologues, other than isozymes, could be identified by sequence analysis (23). It was initially thought that this distinctiveness might be related to the small monomer size of 4-OT (62 amino acids), which was the smallest reported size for an enzyme subunit at the time. Since then, several homologues have been identified as a result of the numerous genome sequencing projects. These proteins, most of which have unidentified physiological functions, are found throughout the bacterial and archaeal kingdoms. Many of the organisms that harbor 4-OT homologues are not known to degrade aromatic hydrocarbons. Furthermore, these homologues have not been found in operons. Although enzymes in metabolic pathways are not required to be grouped in operons, the presence of one homologue in an operon might provide clues about its physiological purpose.

The *ywhB* gene, from *Bacillus subtilis* 168, showed the highest sequence identity to 4-OT (~36%) of any homologue in 1998, when it was first identified. Its cellular purpose was (and remains) unknown. Interest in this particular protein was heightened upon the discovery of *trans*-3-chloroacrylic acid dehalogenase. YwhB shows a similar amount of sequence identity (~35%) to the α-subunit of CaaD (14), raising the question of whether the *ywhB* gene gave rise to the genes for 4-OT and/or CaaD. To investigate the relationship between YwhB and 4-OT, the *Bacillus* enzyme was overexpressed, purified, and characterized with a number of potential substrates. Five mutants were constructed and also characterized. These mutants were chosen on the basis of sequence alignment and the crystal structures of the two enzymes. Both 4-OT and YwhB share Pro-1 and Arg-11, which are necessary for binding and catalysis. Arg-39, which is required for both binding and catalysis in 4-OT, is missing in YwhB, and is replaced with a valine. Lys-37 of YwhB may provide an equivalent charge in the active site. Thus, the P1A and R11A mutants were constructed to determine their relative importance in catalysis. The K37A, V39R,

and K37A V39R mutants were constructed to mimic the potential evolution of YwhB into 4-OT.

Our intensive kinetic studies of YwhB indicate that it although it is an efficient tautomerase ($k_{cat}/K_m$ values ranging from $10^5$-$10^6$ $M^{-1}$ $s^{-1}$), it is only a modest isomerase ($k_{cat}/K_m$ $10^4$ $M^{-1}$ $s^{-1}$). The compounds used to test YwhB for activity were primarily chosen on the basis of the substrates for other members of the tautomerase superfamily. 4-OT readily tautomerizes **2** to **1** and to **3**, and it tautomerizes **5** with a lower efficiency. Another member of the superfamily, macrophage migration inhibitory factor (MIF), preferentially ketonizes **7** and **9**. 5-(Carboxymethyl)-2-oxo-3-hexene-1,6-dioate decarboxylase (COHED), which is not a member of the tautomerase superfamily, tautomerizes **11** in the process of its decarboxylative reaction (16). It can be reasonably inferred that the preferred substrate, among those investigated, is **5**. However, in the absence of a crystal structure or stereochemical analyses, the relative preference YwhB has for these substrates cannot be explained. The compounds examined carry a wide range of functional groups that may be bound in different orientations by the enzyme. Both stereochemical and crystallographic investigations are underway.

The mutations of P1A and R11A in YwhB were performed to verify the importance of each residue in catalysis. The corresponding mutations in 4-OT have been performed and the mutant enzymes characterized, providing a useful reference point. As expected, both are required for catalysis. Of the four reactions investigated (Tables 2-5) only the reaction of **7** was observable. Both mutants showed large drops in $k_{cat}$ (100-fold and 2300-fold, respectively), although the $K_m$ values also dropped. The drop in $K_m$ for the R11A mutant was unexpected. An increase in the $K_m$, as observed for the corresponding 4-OT mutant, would have been consistent with its proposed role in binding. Determination of activity for the other three reactions was precluded by the large amounts of enzyme required. Further investigation of these mutants with HPD as the substrate may provide more information as to their precise roles, particularly in the case of R11A YwhB.

The K37A mutation of YwhB was performed to see if the lysine residue performed a role similar to that of arginine-39 in 4-OT. Both residues are positively charged under physiological conditions. As mentioned previously, lysine is likely not able to adequately replace arginine because it cannot form two hydrogen bonds in the same fashion. The same reactions were investigated as for the other mutants. The decreases in tautomerization ability for **7** and **9** were 10-20-fold for $k_{cat}$ and $K_m$. The largest decreases in activities were observed with **2** as substrate. The isomerization ability was eliminated. The 1,3-keto-enol tautomerization ability dropped 50-fold in $k_{cat}$ and ~100-fold in $k_{cat}/K_m$, while the value for $K_m$ increased two-fold. This suggests that the lysine may play more of a role in catalysis than in binding. Lys-37 may supply a hydrogen bond to C-2 of **2**, facilitating deprotonation at C-3.

Both the V39R and the K37A V39R mutants were made with the intent of evolving YwhB into a better 4-OT. Initially the V39R mutant was made to position the positively charged arginine residue for optimal catalysis of the 1,3- and 1,5-keto-enol tautomerizations. The V39R mutant showed the best ability to catalyze each of the reactions investigated. Most notably, the tautomerization and isomerization abilities improved by 10-fold and 5-fold in terms of $k_{cat}/K_m$, respectively. Thus, the mutation of valine-39 to an arginine may have been one step involved in the evolution of 4-OT from YwhB.

The parameters observed for V39R YwhB led to the design of the K37A V39R double mutant. In V39R YwhB, the presence of Lys-37 results in two positive charges in close proximity, which may be detrimental to catalysis. Thus elimination of K37A would eliminate one of the neighboring positive charges in the V39R mutant. It was presumed that the V39R mutation would continue to fill the role apparently observed in the single mutant. Unfortunately, the K37A V39R double mutant failed to yield a better isomerase. In fact, the double mutations diminished all of the observed activities in terms of $k_{cat}/K_M$. Thus, it is impossible that these two mutations alone gave rise to 4-OT from YwhB as a progenitor.

The results of this study indicate the importance of active-site residues identified by sequence alignment and crystallographic studies of YwhB. Common methods for determining the evolution of proteins include random mutagenesis and directed evolution. However, in the absence of a simple screening technique for mutants, these methods lose their usefulness. In the case of YwhB and 4-OT, the crystal structures have been solved, allowing specific residues to be chosen for mutagenesis experiments. Pro-1 and Arg-11 are critical for the activity of YwhB, as expected from the mechanism of 4-OT. The differences in activity observed for the P1A and R11A YwhB mutants with **7** as substrate are similar to those observed for the P1A and R11A 4-OT mutants with **2** as substrate (9, 22). Lys-37 appears to function in catalysis, perhaps fulfilling the role vacated by the "missing" arginine at position 39. Mutation of Val-39 to an arginine increases the catalytic efficiency of YwhB while maintaining its broad substrate specificity. Interestingly, removing Lys-37 in the presence of Arg-39 is detrimental to its activity. The presence of multiple charged groups in the active site pocket thus seems to be important. Perhaps with the addition of arginine-39, lysine-37 is able to bind an ordered water molecule as is observed in 4-OT. Crystallographic and stereochemical studies, along with the investigation of a K37S V39R mutant, may further elucidate the roles of these residues in catalysis and indicate the evolutionary steps taken between these enzymes.

**ACKNOWLEDGMENTS.**

Electrospray ionization mass spectrometry was performed by the analytical instrumentation service core supported by Center grant ES 07784. We would like to thank Dr. Herng-Hsiang Lo (Division of Pharmacology and Toxicology, The University of Texas at Austin) for obtaining mass spectra, Dr. William H. Johnson, Jr. (Division of Medicinal Chemistry, The University of Texas at Austin) for the synthesis of needed compounds, and Dr. Robert M. Czerwinski (Division of Medicinal Chemistry, The University of Texas at Austin; Genetics Institute) for cloning and protein purification.

**REFERENCES**


1. Whitman, C. P., *et al. JACS,* 1992, **114**, 10104-10110.
2. Fitzgerald, M. C., Chernushevich, I., Standing, K. G., Kent, S. B. H., and Whitman, C. P., *JACS*, 1995, **117**, 11075-11080.
3. Stivers, J. T., *et al*. *Biochemistry*, 1996, **35**, 803-813.
4. Johnson, W. H. Jr., Czerwinski, R. M., Fitzgerald, M. C., and Whitman, C. P. *Biochemistry*, 1997, **36**, 15724-15732.
5. Subramanya, H. S., *et al. Biochemistry*, 1996, **35**, 792-802.
6. Taylor, A. B., Czerwinski, R. M., Johnson, W. H. Jr., Whitman, C. P., and Hackert, M. L. *Biochemistry*, 1998, **37**, 14692-14700.
7. Fitzgerald, M. C., Chernushevich, I., Standing, K. G., Whitman, C. P., and Kent, S. B. H., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 6851-6856.
8. Stivers, J. T., *et al. Biochemistry*, 1996, **35**, 814-823.
9. Harris, T. K., *et al. Biochemistry*, 1999, **38**, 12343-12357.
10. Czerwinski, R. M., Harris, T. K., Massiah, M. A., Mildvan, A. S., and Whitman, C. P., *Biochemistry*, 2001, **40**, 1984-1995.
11. Murzin, A. G. *Curr. Opin. Struct. Biol.*, **6**, 386-394.
12. Whitman, C. P., *Arch. Biochem. Biophys.*, 2002, **402**, 1-13.
13. Almrud, J. J., *et al. Biochemistry* 2002, **41**, 12010-12024.
14. Poelarends, G. J., Saunier, R., and Janssen, D. B. *J. Bacteriol.,* 2001, **183**, 4269-4277.
15. Whitman, C. P., Aird, B. A., Gillespie, W. R., and Stolowich, N. J. *JACS,* 1991, **113**, 3154-3162.
16. Johnson, W. H., Jr., Hajipour, G., and Whitman, C. P. *JACS*, 1995, **117**, 8719-8726.
17. Sambrook, J., Fritsch, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual, 1989.
18. Laemmli, U. K. *Nature*, 1970, **227**, 680-685.

19. Waddell, W. J. *J. Lab. Clin. Med.,* 1956, **48**, 311-314.

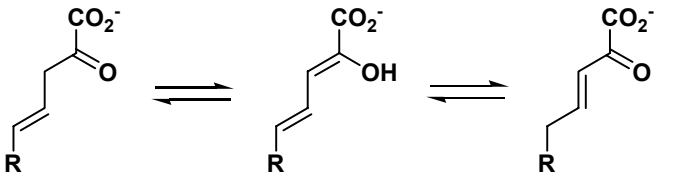
20. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. *Gene,* 1989, **77**, 51-59.

21. Johnson, W. H., Jr., Czerwinski, R. M., Stamps, S. L., and Whitman, C. P. *Biochemistry,* 1999, **38**, 16024-16033.

22. Czerwinski, R. M., *et al. Biochemistry*, 1997, **36**, 14551-14560.

23. Chen, L. H., *et al. J. Biol. Chem.*, 1992, **267**, 17716-17721.

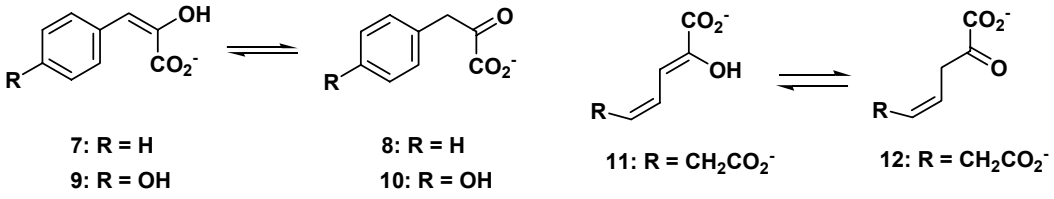
**Scheme 22**




**1: R = $CO_2^-$** **2: R = $CO_2^-$** **3: R = $CO_2^-$**

**4: R = H** **5: R = H** **6: R = H**

**Scheme 23**




**7: R = H**
**9: R = OH**

**8: R = H**
**10: R = OH**

**11: R = $CH_2CO_2^-$**

**12: R = $CH_2CO_2^-$**

**Table 6. Kinetic Parameters for YwhB[a]**

| reaction | $k_{cat}$ ($s^{-1}$) | $K_m$ ($\mu M$) | $k_{cat}/K_m$ ($M^{-1}s^{-1}$) |
|---|---|---|---|
| **2 → 1** | 10.1 ± 1.1 | 75.9 ± 17.0 | $1.3 \times 10^5$ |
| **5 → 4** | 177.4 ± 10.3 | 176.2 ± 19.0 | $1.0 \times 10^6$ |
| **7 → 8** | 124.8 ± 18.2 | 303.3 ± 58.9 | $4.1 \times 10^5$ |
| **9 → 10** | 4.1 ± 0.4 | 157.1 ± 21.8 | $2.6 \times 10^4$ |
| **11 → 12** | 3.0 ± 0.4 | 334.8 ± 55.0 | $9.0 \times 10^3$ |
| **2 → 3** | 26.0 ± 1.4 | 938.9 ± 69.6 | $2.8 \times 10^4$ |

[a]The steady-state kinetic parameters were determined in 20 mM sodium phosphate buffer (pH 7.3) at 23 °C. Errors are standard deviations**.**

**Table 7. Kinetic Parameters for YwhB and Its Mutants: Tautomerization of 2-HM (2)[a]**

| enzyme | $k_{cat}$ ($s^{-1}$) | $K_m$ (μM) | $k_{cat}/K_m$ ($M^{-1}$ $s^{-1}$) | relative $k_{cat}$ | relative $k_{cat}/K_m$ |
|---|---|---|---|---|---|
| wild type | 10 ± 1 | 76 ± 17 | $1.3 \times 10^5$ | 1 | 1 |
| P1A | n/a[b] | n/a | n/a | n/a | n/a |
| R11A | n/a[b] | n/a | n/a | n/a | n/a |
| K37A | 0.2 ± 0.02 | 133 ± 33 | $1.2 \times 10^3$ | 0.02 | 0.0092 |
| V39R | 211 ± 30 | 161 ± 35 | $1.3 \times 10^6$ | 20.9 | 10 |
| K37A V39R | 21 ± 4 | 268 ± 60 | $7.8 \times 10^4$ | 2.1 | 0.6 |

[a]The steady-state kinetic parameters were determined in 20 mM sodium phosphate buffer (pH 7.3) at 23 °C. Errors are standard deviations**.**

[b]Activity was not detectable.

**Table 8. Kinetic Parameters for YwhB and Its Mutants: Isomerization of 2-HM (2)[a]**

| enzyme | $k_{cat}$ ($s^{-1}$) | $K_m$ ($\mu M$) | $k_{cat}/K_m$ ($M^{-1}$ $s^{-1}$) | relative $k_{cat}$ | relative $k_{cat}/K_m$ |
|---|---|---|---|---|---|
| wild type | 26 ± 1 | 939 ± 70 | $2.8 \times 10^4$ | 1 | 1 |
| P1A | n/a[b] | n/a | n/a | n/a | n/a |
| R11A | n/a[b] | n/a | n/a | n/a | n/a |
| K37A | n/a[b] | n/a | n/a | n/a | n/a |
| V39R | 78.3 ± 3.9 | 545 ± 46 | $1.4 \times 10^5$ | 3.01 | 5 |
| K37A V39R | 5.6 ± 1.7 | 2748 ± 930 | $2.0 \times 10^3$ | 0.21 | 0.071 |

[a]The steady-state kinetic parameters were determined in 20 mM sodium phosphate buffer (pH 7.3) at 23 °C. Errors are standard deviations**.**

[b]Activity was not detectable.

**Table 9. Kinetic Parameters for YwhB and Its Mutants: Tautomerization of Phenylpyruvate (7)[a]**

| enzyme | $k_{cat}$ ($s^{-1}$) | $K_m$ (μM) | $k_{cat}/K_m$ ($M^{-1}$ $s^{-1}$) | relative $k_{cat}$ | relative $k_{cat}/K_m$ |
|---|---|---|---|---|---|
| wild type | 125 ± 18 | 303 ± 59 | $4.1 \times 10^5$ | 1 | 1 |
| P1A | 1.2 ± 0.1 | 208 ± 24 | $5.8 \times 10^3$ | 0.0096 | 0.014 |
| R11A | 0.055 ± 0.003 | 82 ± 9 | 669 | $4.4 \times 10^{-4}$ | 0.0016 |
| K37A | 10 ± 2 | 217 ± 50 | $4.8 \times 10^4$ | 0.08 | 0.12 |
| V39R | 48 ± 3 | 193 ± 21 | $2.5 \times 10^5$ | 0.38 | 0.61 |
| K37A V39R | 2.6 ± 0.1 | 178 ± 16 | $1.5 \times 10^4$ | 0.021 | 0.037 |

[a]The steady-state kinetic parameters were determined in 20 mM sodium phosphate buffer (pH 7.3) at 23 °C. Errors are standard deviations.

**Table 10. Kinetic Parameters for YwhB and Its Mutants: Tautomerization of p-Hydroxyphenylpyruvate (9)[a]**

| enzyme | $k_{cat}$ ($s^{-1}$) | $K_m$ (μM) | $k_{cat}/K_m$ ($M^{-1}$ $s^{-1}$) | relative $k_{cat}$ | relative $k_{cat}/K_m$ |
|---|---|---|---|---|---|
| wild type | 4.1 ± 0.4 | 157 ± 22 | $2.6 \times 10^4$ | 1 | 1 |
| P1A | n/a[b] | n/a | n/a | n/a | n/a |
| R11A | n/a[b] | n/a | n/a | n/a | n/a |
| K37A | 0.7 ± 0.1 | 195 ± 40 | $3.4 \times 10^3$ | 0.17 | 0.13 |
| V39R | 4.8 ± 0.4 | 264 ± 31 | $1.8 \times 10^4$ | 1.17 | 0.69 |
| K37A V39R | n/a[b] | n/a | n/a | n/a | n/a |

[a]The steady-state kinetic parameters were determined in 20 mM sodium phosphate buffer (pH 7.3) at 23 °C. Errors are standard deviations**.**

[b]Activity was not detectable.

**FOOTNOTES**

[1]Abbreviations: Cm, chloramphenicol; CaaD, *trans*-3-chloroacrylic acid dehalogenase; ESI-MS, electrospray ionization mass spectrometry; HPLC, high performance liquid chromatography; IPTG, isopropyl-β-D-thiogalactoside; Kn, kanamycin; LB, Luria-Bertani Medium; NMR, nuclear magnetic resonance; 4-OT, 4-oxalocrotonate tautomerase; PCR, polymerase chain reaction; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

# REACTIONS OF 4-OXALOCROTONATE TAUTOMERASE AND YWHB WITH *TRANS*-3-HALOACRYLATES AND 3-HALOPROPIOLIC ACIDS: ANALYSIS AND IMPLICATIONS

## INTRODUCTION

4-Oxalocrotonate tautomerase (4-OT, EC 5.3.2), a bacterial isomerase found on the TOL plasmid in *Pseudomonas putida mt-2* (1), catalyzes the conversion of 2-oxo-4-hexenedioate (**1**, Scheme 24) to 2-oxo-3-hexenedioate (**3**) through a dienol intermediate, known commonly as 2-hydroxymuconate (**2**). The presence of the TOL plasmid enables this soil bacterium to use aromatic hydrocarbons like toluene as its sole sources of carbon and energy. One of the more salient features of 4-OT is the small monomers (62 amino acids) that comprise its hexamer. The enzyme has been intensively studied over the past 15 years by a battery of techniques, primarily due to the simplicity of the reaction coupled with the small monomer size (2-9).

As a result of these efforts, the basic elements of the mechanism have been established. The amino-terminal proline functions as a general base catalyst. Arg-39 and an ordered water molecule provide hydrogen bonds to the carbonyl oxygen of **1**, thereby facilitating deprotonation at C-3 (10). Arg-11 has a binding and a catalytic role: it interacts with the C-6 carboxylate group and functions as an electron sink to draw electron density to the C-5 position for protonation. Phe-50 plays a structural role in maintaining the hydrophobic active site environment, which assists in catalysis and contributes to the lowered p$K_a$ of Pro-1 (11).

Despite these extensive mechanistic studies, it was only with the recent discovery of *trans*-3-chloroacrylic acid dehalogenase (CaaD), that the structural and mechanistic versatility of the β–α–β structural motif encoded by the 4-OT monomer was realized. CaaD transforms the *trans*-isomers of 3-bromo- and 3-chloroacrylates (**4** and **5**, respectively, Scheme 25) to malonate semialdehyde (**7**), presumably through an unstable halohydrin intermediate (**6**) (12). CaaD is found in a degradative pathway

for 1,3-dichloropropene (13), which is used as a nematocide. The enzyme is a heterohexamer comprised of three α -subunits and three β-subunits.

Sequence analysis and mutagenesis identified CaaD as a member of the 4-OT subfamily of the tautomerase superfamily (12). The α-subunit of CaaD shares 35% sequence identity with YwhB, a 4-OT homologue found in *Bacillus subtilis*, while the β-subunit has 25% sequence identity with 4-OT. Both subunits have an amino-terminal proline, one of which, the β-Pro-1, is a critical mechanistic residue that functions as a general acid catalyst to protonate C-2. Arginine-11 in the α-subunit is also important in the mechanism (11).

In view of these sequence and mechanistic similarities, we examined 4-OT and YwhB for low-level dehalogenase activity and found that both exhibited the activity, resulting in the dehalogenation of both **4** and **5**. The presence of this activity further substantiated a relationship between CaaD and the 4-OT family members, 4-OT and YwhB. It also prompted us to examine the behavior of 4-OT and YwhB with 3-bromo- and 3-chloropropiolic acid (**10** and **11**, Scheme 26). CaaD converts the 3-halopropiolic acids into highly reactive acyl halides, which inactivate the enzyme by covalent modification at the β-Pro-1 (14). A similar result with 4-OT and/or YwhB would provide further evidence for the hydratase activity.

Incubation of the enzymes with these two compounds results in their irreversible inactivation. However, the rates of inactivation are much faster (i.e., minutes) than would be anticipated on the basis of the observed rates for the low level hydratase activity (i.e, several days), suggesting that another mechanism is operative. Herein, we report the characterization of the inactivation process for both enzymes, the identification of the covalently modified residue, Pro-1, and propose a mechanism leading to the modification. The mechanism is different from that proposed for the inactivation of CaaD, highlighting differences between the active sites. Although irreversible inhibitors of 4-OT have previously been reported (5, 6), **10** and **11** are the first reported irreversible inhibitors of YwhB. Moreover, the inactivation of YwhB by the modification of its Pro-1 confirms the importance of this residue in the

reaction. These compounds will also be useful as ligands for crystallographic studies to further delineate the roles of active site residues in the YwhB-catalyzed reaction.

## MATERIALS AND METHODS

*Materials*. All reagents, buffers, and solvents were obtained from Sigma Aldrich Chemical Co. (St. Louis, MO), Fisher Scientific Inc. (Pittsburgh, PA), Spectrum Laboratory Products, Inc. (New Brunswick, NJ), or EM Science (Cincinnati, OH), unless noted otherwise. The Amicon stirred cells, ultrafiltration membranes, and Ultrafree-15 centrifugal concentrators were obtained from Millipore Corp. (Billerica, MA). 4-OT, YwhB, P1A- and R11A-4-OT, and P1A- and R11A-YwhB were purified by previously described procedures (6). Synthetic 4-OT was a gift from Dr. Michael C. Fitzgerald (Department of Chemistry, Duke University). Literature procedures were used for the syntheses of *trans*-3-bromoacrylate (**4**), and 3-bromo- and 3-chloropropiolic acids (**10** and **11**) (15, 16). The *trans*-3-chloroacrylate (**5**) was obtained from Fluka Chemical Corp. (Milwaukee, WI). Sequencing-grade trypsin was from Promega Corp. (Madison, WI).

*Methods*. Kinetic data were obtained on an Agilent 8453 Diode Array spectrophotometer. The solutions in the cuvettes were mixed using a stir/add cuvette mixer (Bel-Art Products, Pequannock, NJ). The kinetic data were fitted by nonlinear regression data analysis using the Grafit program (Erithacus Software Ltd., Horley, U. K.) obtained from Sigma Chemical Co. HPLC was performed on a Beckman System Gold HPLC (Fullerton, CA) using a TSKgel Phenyl-5PW hydrophobic column (Tosoh Bioscience, Montgomeryville, PA). Protein was analyzed by Tris glycine sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions on 17.5% gels using either the Mini-Protean II or III vertical gel electrophoresis apparatus obtained from Bio-Rad (Hercules, CA) (17). Protein concentrations were determined using the method of Waddell (18). The NMR spectra were recorded in 100% $H_2O$ on a Varian Unity INOVA-500 spectrometer using selective pre-saturation of the water signal with a 2-s pre-saturation interval. The lock signal is dimethyl-$d_6$ sulfoxide. Chemical shifts are standardized to the dimethyl-$d_6$ sulfoxide signal at 2.49 ppm.

*Enzyme Assays*. The activity of 4-OT was determined by a previously described assay using **2** (19), following the increase in absorbance at 236 nm due to the formation of **3**. The activity of YwhB was determined by a previously described assay using phenylpyruvate (**12**). Ketonization of **12** to generate **13** is followed by the decay at 288 nm (20).

*The 4-OT-catalyzed Hydration of trans-**5***. To an NMR tube containing 100 mM $Na_2HPO_4$ buffer (0.6 mL, pH 9.2) was a added solution of *trans*-**5** (4 mg, 0.04 mmol) dissolved in DMSO-$d_6$ (30 μL). The addition of *trans*-**5** adjusted the pH of the buffer to 6.8. Subsequently, an aliquot of 4-OT (50 μL of a 12.7 mg/mL solution) was added to the NMR tube to initiate the reaction. After incubating at room temperature for 136 h, a $^1$H NMR spectrum was obtained of the reaction mixture was recorded. The $^1$H NMR analysis indicated the presence of **5** (~26%), **8** (~38%), and **9** (~36%). **5**: $^1$H NMR ($H_2O$, 500 MHz) δ 6.09 (1H, d, $J$ = 13 Hz, H2), 6.89 (1H, d, $J$ = 13.5 Hz, H3). **8**: $^1$H NMR ($H_2O$, 500 MHz) δ 2.04 (3H, d, $J$ = 13 Hz, H2), 9.48 (1H, q, H1). **9**: $^1$H NMR ($H_2O$, 500 MHz) δ 1.13 (3H, d, H2), 5.5 (1H, q, H1). The relative amounts of **5**, **8**, and **9** were determined by integration of the C-3 proton of **5**, the C-1 proton of **8**, and the C-2 proton of **9**, respectively.

The experiment was repeated twice. Incubation of 4-OT (50 μL of a 12.7 mg/mL solution) with *trans*-**5** for 96 hr resulted in 71% conversion of **5** to **8** and **9**. Incubation of 4-OT (9 μL of a 21.5 mg/mL solution) with *trans*-**5** for 139 hr resulted in 64 % conversion of **5** to **8** and **9**.

Incubation of *trans*-**5** under the same conditions (for 208 h) in the absence of enzyme resulted in no detectable amount of **8** or **9**.

*Incubation of P1A- and R11A-4-OT with trans-**5***. The two reactions were made up in NMR tubes as described above for the wild type enzyme. The P1A-4-OT-catalyzed reaction was initiated by the addition of 60 μL of a 12.4 mg/mL solution of enzyme. The final pH of the solution was 6.8. The R11A-4-OT-catalyzed reaction was initiated by the addition of 25 μL of a 26.9 mg/mL solution of enzyme. The final pH of the solution was 6.8. After 114 hr at room temperature, the P1A

reaction showed ~2% conversion to **8** and **9**, while the R11A mutant converted ~74% of **5** to **8** and **9**. A second experiment involving the incubation of R11A (12 μL of a 17.7 mg/mL solution) with *trans*-**5** for 208 hr resulted in ~49 % conversion of **5** to **8** and **9**.

*The 4-OT-catalyzed Hydration of trans-**4***. Incubation of 4-OT (20 μL of a 21.5 mg/mL solution) with *trans*-**4** for 209 hr resulted in ~92% conversion of **5** to **8** and **9**. The incubation of *trans*-**4** under the same conditions (for 552 h) in the absence of enzyme resulted in no detectable amount of **8** or **9**.

*Incubation of 4-OT with cis-**5***. Incubation of 4-OT (50 μL of a 12.7 mg/mL solution) with *cis*-**5** for 139 hr resulted in ~3% conversion of **5** to **8** and **9**. Incubation of 4-OT (20 μL of a 21.5 mg/mL solution) with *cis*-(**5**) for 209 hr resulted in ~4% conversion of **5** to **8** and **9**.

*The Synthetic 4-OT-catalyzed Hydration of trans-**5***. Two experiments were carried out. Incubation of synthetic 4-OT (300 μL of a 1.2 mg/mL solution) with *trans*-**5** for 208 hr resulted in ~67% conversion of **5** to **8** and **9**. Repeating the experiment with a different amount of synthetic 4-OT (100 μL of a 2 mg/mL solution) for 208 hr resulted in ~22% conversion of **5** to **8** and **9**.

*The YwhB-catalyzed Hydrations of trans-**4** and **5***. The NMR experiments were set up as described above. Incubation of YwhB (50 μL of a 12.7 mg/mL solution) with *trans*-**4** for 616 hr resulted in ~93% conversion of **5** to **8** and **9.** Incubation of the same amount of YwhB *trans*-(**5**) for 171 hr resulted in ~30% conversion of **5** to **8** and **9.**

*Incubation of P1A- and R11A-YwhB with trans-**5***. The two reactions were made up in NMR tubes as described above for the wild type enzyme. The P1A-YwhB-catalyzed reaction was initiated by the addition of 100 μL of a 5.4 mg/mL solution of enzyme. The R11A-YwhB-catalyzed reaction was initiated by the addition of 50 μL of an 11.9 mg/mL solution of enzyme. After 108 h, there was no detectable product in the reaction mixture containing the P1A-YwhB mutant. After 137 h, the R11A-YwhB mutant had converted ~5% of *trans*-**5** to **8** and **9**.

*Incubation of YwhB with cis-**5***. Incubation of YwhB (50 μL of a 6.6 mg/mL solution or 20 μL of a 25.6 mg/mL solution) with *cis*-**5** for 137 hr and 209 h, respectively, resulted in no detectable conversion of *cis*-**5** to **8** and **9**.

*Incubation of pET-24a(+) and Fumarase with trans-**5***. An aliquot of pET-24a(+) (0.5 μL) was transformed into *E. coli* strain BL21-Gold(DE3)pLysS by electroporation following the manufacturer's directions. Cells carrying the pET-24a(+) plasmid were grown and induced for protein expression as described for the production of YwhB and 4-OT. Cells were stored at -80° C until use. Cultures (500 mL) grown under these conditions yield 1.34 g of cells.

In a modified partial purification, the cells were re-suspended in 10 mL of 10 mM ethylenediamine buffer (pH 7.4), placed on ice, and initially disrupted by sonication at 50% output for 15 s. Phenylmethylsulfonyl fluoride (0.5 mM) and 6-aminocaproic acid (1 mM) were added to inhibit cellular proteases. Subsequently, the cell suspension was subjected to further sonication for two 5-min intervals using 5-s pulses with a cycle time of 50%. Between intervals, the cells were allowed to sit for 5 minutes. The lysate was centrifuged at 19,000 rpm for 30 minutes at 4 °C. The resulting supernatant was made 2M in $(NH_4)_2SO_4$ and stirred at 4 °C for 2 hr. Subsequently, the solution was centrifuged for 30 minutes at 19,000 rpm. The supernatant was transferred to an Ultrafree 15 (MWCO 5000 Da) centrifugal concentrator and exchanged into 20 mM sodium phosphate buffer (pH 7.3) to a final volume of 0.6 mL. The concentration of the protein was estimated by the Waddell method to be 131.3 mg/mL.

Incubation of the partially purified protein solution generated from the empty pET vector (20 μL of a 1331.3 mg/mL solution) with either *trans*-**4** or *trans*-**5** for 252 hr resulted in no detectable amount of **8** or **9**.

Incubation of fumarase (100 μL of a 0.5 mg/mL solution) with *trans*-**5** for 112 hr resulted in no detectable amount of **8** or **9**.

*Irreversible Inhibition of 4-OT by **10** and **11***. The inactivation of 4-OT by **10** or **11** was determined by incubating various concentrations of inhibitor (**10**: 50-600

μM and **11**: 50-750 μM) with 4-OT in 20 mM sodium phosphate buffer (pH 6.0) at 30°C. At each inhibitor concentration, three runs were performed. Stock solutions of **10** and **11** (50 mM) were made in 100 mM $Na_2HPO_4$ buffer and had a final pH ~7. The 50 mM stock solutions were diluted with a sufficient volume of the 100 mM $NaH_2PO_4$ buffer (pH 7.3) to give 5 mM solutions of **10** or **11**. A quantity of 4-OT (4 μL of a 21.5 mg/mL solution) was diluted 2000-fold into 20 mM sodium phosphate buffer (pH 6.0), resulting in a final concentration of 1.58 μM. The diluted enzyme was stored at 4 °C. Subsequently, the enzyme was divided into 105 μL aliquots and equilibrated at 30 °C for ~5 min prior to use. The time zero point (i.e., 100% activity) was determined at each inhibitor concentration by removing and assaying an aliquot (5 μL) of enzyme just before the addition of inhibitor. After the addition of inhibitor, aliquots (5 μL) were removed at 7-s intervals, (for a total 56 s), diluted into 1 mL of 20 mM $NaH_2PO_4$ buffer (pH 7.3), and assayed for residual 4-OT activity. The assay was initiated by the addition of **2** to give a final concentration of 500 μM. The volume of inhibitor added did not exceed 7.5% of the total volume of the incubation mix.

*Protection of 4-OT from Inactivation by **10** Using **1-3***. The protection studies of 4-OT from inactivation by **10** were carried out using the equilibrium mixture of **1-3** as described elsewhere (6), with the following modifications. A stock solution of **2** (83.5 mM) was made in 100 mM $Na_2HPO_4$ buffer (pH 9.2), resulting in a final pH of ~6.8. The stock solution was diluted 10-fold into 100 mM $NaH_2PO_4$ buffer (pH 7.3). Quantities of 4-OT (105 μL) were prepared as described above. Aliquots of **2** (0-5 μL from the 8.35 mM solution resulting in concentrations ranging from 0-1.04 mM) were added to the enzyme and incubated at 30 °C for ~5 min, thereby allowing the mixture of **1-3** to reach equilibrium. Subsequently, an aliquot of **10** (2.5 μL from a 5 mM solution) was added, resulting in a final concentration of 125 μM. Aliquots (5 μL) were removed at 7-s intervals (for a total of 56 s), diluted 200-fold into 20 mM sodium phosphate buffer (1 mL, pH 7.3) and assayed for residual activity as described above.

*Irreversibility of the Inactivation of 4-OT*. 4-OT (1.58 μM based on the molecular weight of the monomer) was incubated with an excess of **10** or **11** (50 μM) in 4 mL of 20 mM $NaH_2PO_4$ buffer (pH 7.3) for 2 hr at 29 °C. The final pH of the solution was ~7. In a separate control reaction, an identical quantity of 4-OT was incubated without inhibitor under identical conditions. The treated samples had no activity after 2 hr. The three samples were dialyzed against 20 mM $NaH_2PO_4$ buffer, pH 7.3 at 4 °C. After 2 weeks, the control sample lost none of its original activity, while the treated samples did not regain any activity.

*Electrospray Ionization (ESI) Mass Spectrometric Analysis of 4-OT and 4-OT Modified by **10** or **11***. Each sample contained ~1 mg 4-OT (47 μL of a 21.5 mg/mL solution) in a final volume of 1 mL of 20 mM sodium phosphate buffer (pH 7.3). The modified samples were generated by the incubation of 4-OT with either **10** or **11** (53 μL of a 50 mM stock solution to give a final concentration of 2.65 mM). The control samples were treated with an aliquot (53 μL) of a 100 mM sodium phosphate buffer (pH 7.3). Samples were incubated for ~18 hr at 4 °C and assayed for activity. The control retained activity while the treated samples had no activity. Subsequently, the samples were passed through separate PD-10 Sephadex columns equilibrated in de-ionized water. Fractions (~0.5 mL) were collected, and protein was identified by the Waddell method (18). Samples were loaded as 50% infusions with methanol (8-15 μL) or through a peptide trap for ESI-MS analysis.

*Proteolytic Digestion of 4-OT and 4-OT modified by **10***. Two samples were made up for mass spectral analysis. Each sample contained ~1 mg of 4-OT (0.5 mL of a 21.5 mg/mL solution) in a sufficient quantity of 20 mM sodium phosphate buffer, pH 7.3, to give a final volume of 2.5 mL. One sample was treated with **10** (50 μL from a 50 mM stock solution of **9** in 100 mM $Na_2HPO_4$ buffer). After incubating the samples at 4 °C for ~12 hrs, the individual samples were exchanged (~99.8 %) into 40 mM ammonium acetate buffer, pH 4.0, using an Amicon equipped with a YM-3 membrane. The treated 4-OT gave a yellow color, while the untreated sample remained colorless. Subsequently, both samples were treated with sequencing grade

endoproteinase Glu-C from *Staphylococcus aureus* (protease V8) by a modification of a literature procedure (6). Accordingly, two vials of endoproteinase Glu-C (25 μg) were each reconstituted in 40 mM ammonium acetate buffer (50 μL), pH 4.0. A quantity (50 μL) of the reconstituted protease was added to each sample, and the two samples were incubated at 37 °C overnight. The mixtures were subjected to matrix assisted laser desorption-ionization (MALDI) mass spectral analysis.

*Irreversible Inhibition of YwhB by **10** and **11***. The inhibition studies of YwhB with **10** and **11** were carried out as described above for 4-OT with the following modifications. A quantity of YwhB (100 μL of a 25.6 mg/mL solution) was diluted 100-fold into 20 mM sodium phosphate buffer (pH 6.0), to give a final concentration of 36.5 μM. The enzyme was stored at 4 °C. The diluted enzyme was divided into 220 μL portions and allowed to equilibrate to 30 °C for ~ 5 min just prior to use. The inactivation of YwhB by **10** or **11** was determined by the incubation of inhibitor (**10**: 50-1000 μM and **11**: 50-750 μM) with enzyme (220 μL quantities) in 20 mM sodium phosphate buffer (pH 7.3) at 30 °C. The volume of inhibitor added did not exceed 4 % of the total volume of the inhibition mix. Aliquots (20 μL) were removed at various time points (for **10**: 10-s intervals for 50-150 μM, 7-s intervals for 200-1000 μM and for **11**: 10-s intervals for 50-100 μM, 7-s intervals for 125-750 μM), diluted into 1 mL of 20 mM sodium phosphate buffer (pH 7.3), and assayed for residual activity. The assay was initiated by the addition of **12** to give a final concentration of 150 μM.

*pH Dependence of the Inactivation of YwhB Using **11***. Inactivation experiments were carried out as described above at five different pH values (pH 5.2, 6, 7, 8, and 9) using a quantity of YwhB (10 μL of a 25.6 mg/mL solution) diluted 100-fold into 20 mM sodium phosphate buffer at the indicated pH values. The diluted enzyme solutions were incubated at 4 °C overnight. The concentration of **11** tested at each pH was 250 μM. Time points were taken at 7-s intervals.

*Protection of YwhB from Inactivation by **10** Using **1-3***. The protection of YwhB from inactivation by **10** was carried out using **1-3** as described above in the 4-

OT protection studies with the following modifications. Quantities of YwhB (220 μL portions made up as described above) were placed in 1.5 mL eppendorf tubes. Aliquots (0-6.6 μL) of **2** were added to the enzyme and the resulting solution was equilibrated for ~5 min. The concentrations of **2** in the eppendorf tubes ranged from 0-2.5 mM. Subsequently, an aliquot of **10** (1.2 μL from a 50 mM solution) was added to the mixture, resulting in a final concentration of 300 μM. Aliquots (20 μL) were removed from the mixtures at 7-s intervals, diluted 50-fold into 20 mM sodium phosphate buffer (1 mL, pH 7.3) and assayed for residual activity as described above.

*Irreversibility of the Inactivation of YwhB*. The irreversibility of the reaction was established using both **10** and **11**. YwhB (142 μM based on the monomer molecular weight) was incubated with an excess of **10** or **11** (3 mM) in 1 mL of 20 mM $NaH_2PO_4$ buffer (pH 7.3) for 2 hr at 29 °C. The final pH of the solution was ~7. In a separate control reaction, the same quantity of enzyme was incubated without inhibitor under similar conditions. The samples treated with inhibitors had no activity after 2 hr. The three samples were dialyzed against 20 mM $NaH_2PO_4$ buffer, pH 7.3 at 4 °C. After 8 days, the control sample lost ~6% of its original activity, while the treated samples did not regain any activity.

*ESI-MS Analysis of YwhB and YwhB Modified by **10** or **11***. Each sample contained ~1 mg YwhB (40 μL of a 25.6 mg/mL solution) in a final volume of 1 mL of 20 mM sodium phosphate buffer (pH 7.3). The modified samples were generated by the incubation of YwhB with either **10** or **11** (60 μL from a 50 mM stock solution to give a final concentration of 3 mM). The control samples contained an equivalent amount of 100 mM sodium phosphate (pH 7.3). Samples were incubated for ~14 hr at 4 °C and assayed for activity. The control samples retained activity while the treated samples had no activity. Subsequently, the samples were passed through separate PD-10 Sephadex columns equilibrated in deionized water. Fractions (~0.5 mL) were collected, and protein was identified by the Waddell method (18). Samples (5-10 μL) were loaded through a peptide trap for ESI-MS analysis.

*Proteolytic Digestion of YwhB and YwhB-modified by **10**.* In order to determine the site of covalent modification, it was necessary to treat YwhB and YwhB modified by **10** with protease V8 and trypsin (in separate reactions). For the proteolytic digestion using protease V8, each sample contained ~0.5 mg of YwhB (20 μL of a 25.6 mg/mL solution) in a sufficient quantity of 20 mM sodium phosphate buffer (pH 7.3) to give a final volume of ~0.5 mL. One sample was treated with **10** (30 μL from a 50 mM stock solution in 100 mM $Na_2HPO_4$ buffer). The samples were incubated at 4 °C for ~20 hrs. Subsequently, each sample was exchanged into ammonium acetate buffer, treated with protease V8 using the protocol described above, and subjected to matrix assisted laser desorption-ionization (MALDI) mass spectral analysis. For the proteolytic digestion using trypsin, the YwhB samples were prepared as described above and incubated at 4 °C for ~14 hrs. Subsequently, they were passed through separate PD-10 Sephadex columns equilibrated in 100 mM $(NH_4)_2CO_3$ buffer, pH 8. Fractions (~0.5 mL) were collected, and protein was identified by the Waddell method. Both samples were treated with sequencing grade trypsin according to the manufacturer's instructions and subjected to (MALDI) mass spectral analysis.

*Mass Spectrometry of the Intact and Digested Proteins*. The intact samples (4-OT, 4-OT-modified by **10** or **11**, YwhB, and YwhB-modified by **10** or **11**) were analyzed using an LCQ electrospray ion trap mass spectrometer (ThermoFinnigan, San Jose, CA), housed in the Analytical Instrumentation Facility Core in the College of Pharmacy at the University of Texas at Austin as described previously (14). The protease-digested control and treated samples (using trypsin or protease V8) were analyzed on the delayed extraction Voyager-DE PRO MALDI-TOF instrument (PerSeptive Biosystems, Framingham, MA) using a previously described protocol (21). The ions in the trypsin-digested treated YwhB modified by **10** were subjected to MALDI-PSD analysis using the protocol described elsewhere (14).

## RESULTS

*The 4-OT and YwhB-Catalyzed Hydrations of E-**4** and E-**5**.* The dehalogenase activities of 4-OT and YwhB was examined with the *E*-isomer of **4** as well as the *E*- and *Z*-isomers of **5**. The results are summarized in Tables 11 and 12. The wild type 4-OT and YwhB have comparable dehalogenase activity using *E*-**5** and may have comparable activities using *E*-**4**. The prolonged incubation for YwhB with *E*-**4** precludes a direct comparison. Interestingly, 4-OT will process the *cis*-isomer of **5**, albeit very slowly, while YwhB will not.

Although the preparations of 4-OT and YwhB were highly purified (as assessed by SDS-PAGE), four control experiments were carried out in order to verify that the observed dehalogenase activity is due to 4-OT (or YwhB), and not a contaminating protein. First, the non-enzymatic decay of both *E*-**4** and *E*-**5** were examined in buffer for several days. There were absolutely no detectable amounts of acetaldehyde in either mixture. Second, *E*-**5** was incubated with fumarase, an abundant cellular protein, for several days. Again, the compound was not hydrated by fumarase. Third, *E*-**4** and *E*-**5** were incubated with a partially purified cell extract derived from the expression cells used to generate 4-OT and YwhB. The "empty" pET-24a(+) vector was transformed into cells, which were subjected to lysis and ammonium sulfate precipitation as performed for of 4-OT and YwhB. The resulting protein concentrate was exchanged into phosphate buffer for NMR analysis. Again, there were absolutely no detectable amounts of acetaldehyde in either mixture despite the prolonged incubation (~10.5 days). Finally, a 4-OT constructed by total chemical synthesis was incubated with *E*-**5**. This enzyme was completely untouched by any biological system, such as culture flasks and columns used for protein purification. The synthetic 4-OT also displayed comparable dehalogenase activity, converting ~67% of *E*-**5** to acetaldehyde and the hydrate after 208 h.

Studies with two mutants of each protein (P1A and the R11A) further confirm that the activity is attributable to the proteins and identify the proline as a critical

residue in the mechanism. Our results show that the N-terminal proline is critical for activity. No detectable amounts of acetaldehyde and its hydrate were produced by the P1A mutant of either 4-OT or YwhB. In contrast, Arg-11 is critical to the YwhB-catalyzed dehalogenation of *E*-**5**, but not to the 4-OT-catalyzed dehalogenation of *E*-**5**.

*Time-Dependent Inactivation of 4-OT by **10** and **11***. The initial inactivation experiments of 4-OT by **10** and **11** were carried out at pH 7.3 (data not shown). The data indicated that it was not possible to saturate the enzyme with inhibitor, which prompted us to examine the behavior of 4-OT with these inhibitors under different conditions. In view of the observation that the inactivation of YwhB by **11** is pH-dependent (*vide infra*) and the enzyme approached saturation kinetics at lower pH values, the 4-OT inhibition studies were carried out at pH 6.0.

Accordingly, incubation of 4-OT with either **10** or **11** results in a rapid time-dependent, irreversible inactivation of the enzyme at pH 6.0 (Figures 37 and 38). For **10**, the $k_{obsd}$ values measured in 27 experiments were plotted versus the inhibitor concentration and fit to a rectangular hyperbola (Figure 37B). The values of $k_{inact}$ and $K_I$ values obtained from this plot are 0.08 ± 0.02 $s^{-1}$ and 0.71 ± 0.23 mM, respectively. Treating the $k_{obsd}$ values measured in 21 experiments for **11** (Figures 2A, B) similarly yielded $k_{inact}$ and $K_I$ values of 0.47 ± 0.16 $s^{-1}$ and 1.91 ± 0.85 mM, respectively. The rapid loss of activity at higher concentrations of **10** (> 600 μM) or **11** (> 750 μM) did not allow us to collect sufficiently precise data to obtain inactivation rates, thereby accounting for the large errors in the values of $k_{inact}$ and $K_I$. Nonetheless, inactivation with both inhibitors nears saturation, indicative that a dissociable complex forms between enzyme and **10** (or **11**) before inactivation. This supposition is supported by the observation that the enzyme could be protected from inactivation (by **10**) when incubated with the equilibrium mixture of **1-3** (Figure 37C). At ~1 mM, 4-OT is completely protected from inactivation by **10** (125 μM). In the absence of **1-3**, the enzyme loses ~80% of its activity within ~21 s. Exhaustive dialysis of 4-OT (inactivated by **10** or **11**) does not result in the reactivation of 4-OT,

which indicates that a covalent bond has formed between 4-OT and **10** (or **11**) or a species derived from **10** (or **11**).

*ESI-MS Analysis of 4-OT and 4-OT Modified by **10** or **11***. In order to determine whether the enzyme had been covalently modified by **10** (or **11**), 4-OT was incubated with each inhibitor in separate reactions, and the products were isolated and analyzed by ESI-MS. The molecular mass of 4-OT increases from 6810.0 (± 3 Da) to 6896.0 (± 3 Da) after incubation with **10**, and it increases to 6895.0 (± 3 Da) after incubation with **11**. Hence, both treated proteins are covalently modified by a species having a mass of 85 (± 3 Da). In addition, the spectra displayed another signal at 6852 (± 3 Da), which is due to modification of 4-OT by species having a mass of 42 (± 3 Da).

*Identification of the Modified Amino Acid Residue in 4-OT by Mass Spectrometry*. In order to identify the covalently modified residue, 4-OT was inactivated by **10**, (designated the treated sample), incubated overnight with protease V8, and was analyzed by MALDI-TOF mass spectrometry. A control sample was made up similarly, but **10** was excluded from the mixture. The site of covalent attachment was identified by a comparison of the two spectra (21). In the MALDI-MS spectrum of the control sample, there is a prominent peak at *m/z* 1033.59 ± 0.6 Da, which corresponds to the calculated mass of the N-terminal sequence (PIAQIHILE) (Table 13). Two smaller peaks at 1055.69 Da and 1071.64 Da are also present, and represent the sodium and potassium adducts of this fragment, respectively. In the MALDI spectrum of the treated sample, these peaks are not present. However, there are three new peaks at 1075.73 Da, 1097.71 Da, and 1113.68 Da. These new peaks are consistent with covalent modification of the N-terminal fragment by a species having a mass of 42 Da. The two potential sites for covalent modification in this fragment are Pro-1 and His-6. It can be reasonably concluded that Pro-1 is the site of attachment because the kinetic studies indicate that **10** binds in the active site and His-6 is not in the active site. Finally, a comparison of the MALDI

MS spectra indicates no other fragments are modified, which is consistent with the results of the ESI-MS analysis (showing a single site of attachment).

*Time-Dependent Inactivation of YwhB by **10** and **11***. When the inactivation experiments (using **10** and **11**) were carried out at 7.3, YwhB was rapidly inactivated, which precluded the collection of accurate data for the determination of $k_{obsd}$ values. Hence, the pH dependence of the inactivation process was examined as a function of pH. Because the process slowed at lower pH values, the inactivation studies of YwhB were carried out at pH 6.0. Like 4-OT, the incubation of YwhB with either **10** or **11** results in a rapid time-dependent, irreversible inactivation of the enzyme (Figures 39 and 40). For **10**, the $k_{obsd}$ values measured in nine experiments were plotted versus the inhibitor concentration and fit to a rectangular hyperbola (Figure 39B). The values of $k_{inact}$ and $K_I$ values obtained from this plot are 0.08 ± 0.02 $s^{-1}$ and 1.60 ± 0.47 mM, respectively. Treating $k_{obsd}$ values measured in twelve experiments for **11** similarly yielded $k_{inact}$ and $K_I$ values of 0.11 ± 0.02 $s^{-1}$ and 0.88 ± 0.18 mM, respectively (Figure 4B). Inactivation with both inhibitors approaches saturation, indicative that a dissociable complex forms between enzyme and **10** (or **11**) before inactivation. The results of the protection studies (using a mixture of **1-3**) further support binding at the active site (Figure 39C). At ~2.5 mM, YwhB is partially protected from inactivation by **10** (300 μM). In the absence of **1-3**, the enzyme loses ~ 80% of its activity within 1 min. In the presence of **1-3** (2.5 mM), the enzyme only loses 60% of its activity within 1 min. Exhaustive dialysis of YwhB (inactivated by **10** or **11**) does not result in the reactivation of YwhB, which indicates that a covalent bond has formed between YwhB and **10** (or **11**) or a species derived from **10** (or **11**).

*ESI-MS Analysis of YwhB and YwhB Modified by **10** or **11***. ESI-MS analysis demonstrated that the incubation of YwhB with **10** or **11** resulted in an increase in its molecular mass from 7013.0 (± 3 Da) to 7099.0 (± 3 Da). This observation is consistent with the covalent attachment of a species to YwhB with a mass of 86 (± 3 Da). ESI-MS analysis also showed another peak at 7055 (± 3 Da), which is

consistent with the covalent modification of YwhB by a species having a molecular mass of 42 (± 3 Da).

*Identification of the Modified Amino Acid Residue in YwhB by Mass Spectrometry*. In order to determine the residue modified by **10**, it was necessary to carry out two separate proteolytic digestions using protease V8 and trypsin. The mixtures were analyzed by MALDI-TOF mass spectrometry. The protease V8 digestion of the YwhB and YwhB-modified by **10** generated nine fragments (Table 14). A molecular mass (1079.84 Da) corresponding to the N-terminal fragment (residues 1-9) was observed in the control sample, and was not present in the treated sample. However, a signal appeared at 1121.58 Da, which corresponds to the N-terminal fragment modified by a species with a molecular mass of 42 Da. Unfortunately, the intensity of this signal placed it just above the noise such that a definitive conclusion could not be made. A second signal appeared in the control sample at 2962.49 Da, corresponding to residues 1-25. This signal was not present in the YwhB sample treated with **10** (and digested by V8), but a new signal appeared at 3004.58, which corresponds the N-terminal fragment modified by a species with a molecular mass of 42 Da. A third signal appeared in the control sample at 1901.95 Da, which corresponds to residues 10-25. The same signal also appears in the sample treated with **10**. These observations localize the site of modification of YwhB by **10** to the first nine amino acids of YwhB (PYVTVKMLE).

In order to identify the modified amino acid within this fragment, YwhB and YwhB modified by **10** were subjected to proteolysis by trypsin. The digestion generated eight fragments with the most prominent signal being displayed at 1135.61 Da, corresponding to residues 38-46. The signal was also present in the sample treated with **10**, indicating that these residues had not been modified, consistent with the results obtained for the V8 digestion. A molecular mass (calc. 706.86 Da) corresponding to one expected product of the trypsin digest, the N-terminal fragment (residues 1-6), was not present in the control sample. However, a signal appeared at 748.38 Da, which corresponds to residues 1-6 modified by a species with a molecular

mass of 42 Da. The MALDI-PSD spectrum of this peptide shows a series of b ions, which have a charge on the N-terminal side of the fragmented peptide (Figure 41). Each ion is modified by a mass of +42 Da (except the $b_5$ ion, which is not detected) indicating that the modification is on the N-terminal proline.

## DISCUSSION

Nature may exploit the catalytic promiscuity observed in some enzymes in order to create new enzymes. A catalytically promiscuous enzyme is one that has multiple low level activities in addition to its primary, physiological activity (22). Duplication of the gene followed by a series of mutations to amplify the desired activity generates a new enzyme which has the progenitor's low-level activity as its primary activity. Our results clearly show that the bacterial isomerase, 4-oxalocrotonate tautomerase (4-OT) and a homologue, YwhB, found in *Bacillus subtilis*, have a low-level dehalogenease activity, converting either *trans*-**4** or *trans*-**5** to acetaldehyde and the hydrate.

The most reasonable scenario for the generation of acetaldehyde from *trans*-**4** or *trans*-**5** is shown in Scheme 25. Hydration of **4** or **5** by 4-OT (or YwhB) produces a chlorohydrin species (**6**), which decomposes to malonate semialdehyde (**7**). The non-enzymatic decarboyxlation of **7** yields acetaldehyde (**8**), which is readily hydrated to form **9**. Intermediate **7** is not sufficiently stable to accumulate in the necessary quantities that can be detected by $^{1}$H NMR spectroscopy. This is consistent with the observation that signals for the intermediates are not detected in the course of the experiment.

The demonstration of dehalogenase activity for 4-OT and YwhB is significant for three reasons. First, it suggests that 4-OT and YwhB could be the progenitors for CaaD. Second, the non-enzymatic reaction is a difficult reaction to carry out, but it is clearly accelerated by 4-OT and YwhB. In 0.5 M aqueous NaOH at 60 °C, about 10% of the chloride is removed from *E*-**5** in 24 hr (23). In contrast, 4-OT and YwhB convert ~75% of *E*-**5** to acetaldehyde (and the hydrate) at room temperature and neutral pH after 5-6 days. Third, the active site pocket of 4-OT is a hydrophobic one which is not set up to carry out a hydration reaction. Despite this fact, the enzyme is able to perform a hydration reaction.

Our results also indicate that Pro-1 plays a central role in the mechanism of the 4-OT-catalyzed dehalogenation reaction while Arg-11 does not. In 4-OT, Pro-1 has a p$K_a$ of ~6.4. This indicates that at pH 7.3, about 88% of the proline is the conjugate base form while 12% is the conjugate acid form. As a general base, Pro-1 could activate a water molecule for nucleophilic attack at C-3. While Arg-11 is not important to the 4-OT-catalyzed dehalogenation reaction, there is another arginine residue, Arg-39, which could be involved in binding the carboxylate group of the substrate and pulling electron density away from C-3 to create a partial positive charge at C-3. This charge would make C-3 susceptible to nucleophilic attack by water. In another scenario, Pro-1 could protonate C-2, thereby creating a positive charge at C-3 that enables water to add at the activated carbon.

These results indicate that both Pro-1 and Arg-11 play a critical role in the mechanism of the YwhB-catalyzed dehalogenation reaction. The p$K_a$ of the Pro-1 is not known in YwhB, but in view of the structural similarities between the two enzymes (10), it seems likely that they have comparable p$K_a$ values. Accordingly, Arg-11 could function as an electron sink to pull electron density away from C-3, thereby creating a partial positive charge, which would make C-3 susceptible to nucleophilic attack by water. Pro-1 must also participate in the process and may be involved in the activation of the water.

In order to obtain further evidence for this hydration reaction, the behavior of 4-OT and YwhB with 3-bromo- and 3-chloropropiolic acid was examined (**10** and **11**). Inactivation of these two enzymes by **10** or **11** would provide further evidence for the hydration reaction if the inactivation proceeded through an acyl halide intermediate, as has been proposed for the inactivation of CaaD by these compounds (14). Surprisingly, the inactivation of 4-OT and YwhB was much faster than would have been anticipated based on the rates for the hydration of *E*-**5**. This observation suggested that an alternative mechanism might be involved.

*A priori*, the modification of Pro-1 can occur by a mechanism-based route to form an acyl halide (or a ketene) or by Michael addition of β-Pro-1 to C-3 (Scheme

27). In the first mechanism (Scheme 27A), the enzyme hydrates **10** (or **11**) to generate an unstable species **12**, which will readily undergo a rearrangement to either the acyl halide (**13**) or the ketene (**14**). Due to their reactivity, both species would be lethal to the enzyme. As indicated above, the rapid rate of inactivation and the slow rate of hydrolysis of *E*-**5** argue against this mechanism.

In the second mechanism (Scheme 27B), Pro-1 attacks the C-3 position of **10** (or **11**) in a Michael-type reaction. While conjugate additions to α,β-unsaturated carboxylic acids are not normally known, the inhibitor would presumably be activated by an interaction between the carboxylate group of **10** and a residue on the enzyme (e.g., Arg-11). Subsequent rearrangement would result in the covalent attachment of the propargyl moiety to the enzyme and the loss of the halide ion.

The covalent adduct on Pro-1 in both 4-OT and YwhB has a molecular mass of 85 Da, which is consistent with the attachment of a 3-oxoacetate moiety. Moreover, this adduct, a β-ketoacid, is quite susceptible to facile decarboxylation upon covalent attachment, resulting in an acetyl group (24). This sequence of events could be consistent with either a mechanism-based route or by the rearrangement of the propargyl moiety as shown in Scheme 28. A non-enzymatic rearrangement of the propargyl moiety would result in the formation of an electrophilic allene-type compound (**15**), which would be highly vulnerable to attack by a water molecule. Ketonization of the resulting enol could generate the observed 3-oxoacetate moiety.

We have previously found that the acetylene compound, 2*E*-fluoro-2,4-hexenynoate (**16**) is a potent competitive inhibitor of both 4-OT and YwhB (unpublished data). For 4-OT, the $K_i$ is 0.19 nM while for YwhB, the $K_i$ is 200 μM. These values parallel the lethality of **10** and **11** for 4-OT and YwhB and suggest that acetylene compounds bind at the active site of both enzymes.

**ACKNOWLEDGMENTS**

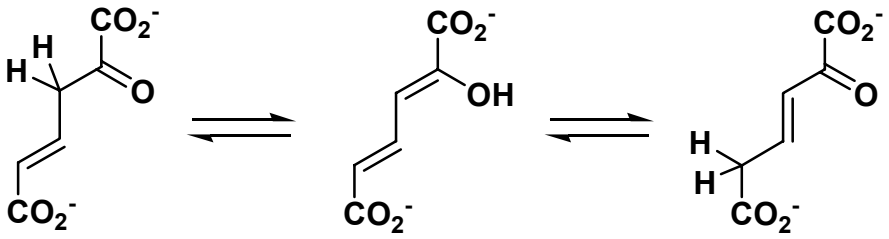
Electrospray ionization (ESI) and matrix assisted laser desorption-ionization (MALDI) mass spectrometry were performed by the analytical instrumentation service core supported by Center grant ES 07784. We thank Steve D. Sorey (Department of Chemistry, The University of Texas at Austin) for his expert assistance in acquiring the NMR spectra. We thank Dr. Wendi M. David and Dr. Maria D. Person (Division of Pharmacology and Toxicology, The University of Texas at Austin) for obtaining the mass spectral data. We thank Dr. William H. Johnson, Jr. (Division of Medicinal Chemistry, The University of Texas at Austin) for the synthesis of needed compounds and obtaining and interpreting NMR spectra. We thank Dr. Michael C. Fitzgerald (Department of Chemistry, Duke University) for the gift of synthetic 4-OT.

**REFERENCES**


1. Horn, J. M., Harayama, S. and Timmis, K. N., *Mol. Microbiol*., 1991, **5**, 2459-2474.
2. Inoue, J., Shaw, J. P., Rekik, M., and Harayama, S. *J Bacteriol.*, 1995, **177**, 1196-201.
3. Whitman, C. P., *et al. JACS,* 1992, **114**, 10104-10110.
4. Fitzgerald, M. C., Chernushevich, I., Standing, K. G., Kent, S. B. H., and Whitman, C. P., *JACS*, 1995, **117**, 11075-11080.
5. Stivers, J. T., *et al*. *Biochemistry*, 1996, **35**, 803-813.
6. Johnson, W. H. Jr., Czerwinski, R. M., Fitzgerald, M. C., and Whitman, C. P. *Biochemistry*, 1997, **36**, 15724-15732.
7. Subramanya, H. S., *et al. Biochemistry*, 1996, **35**, 792-802.
8. Taylor, A. B., Czerwinski, R. M., Johnson, W. H. Jr., Whitman, C. P., and Hackert, M. L. *Biochemistry*, 1998, **37**, 14692-14700.
9. Fitzgerald, M. C., Chernushevich, I., Standing, K. G., Whitman, C. P., and Kent, S. B. H., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 6851-6856.
10. Whitman, C. P., *Arch. Biochem. Biophys.*, 2002, **402**, 1-13.
11. Czerwinski, R. M., *et al. Biochemistry*, 1997, **36**, 14551-14560.
12. Poelarends, G. J., Saunier, R., and Janssen, D. B. *J. Bacteriol.*, 2001, **183**, 4269-4277.
13. Poelarends, G. J., Kulakov, L. A., Larkin, M. J., van Hylckama Vlieg, J. E. T., and Janssen, D. B. *J. Bacteriol.,* 2000, **182**, 2191-2199.
14. Wang, S. C., Person, M. D., Johnson, W. H., Jr., and Whitman, C. P. *Biochemistry*, 2003, **42**, 8762-8773.
15. Strauss, F., Kollek, L., and Heyn, W. *Chem. Ber.*, 1930, **63**, 1868-1899.
16. Andersson, K. *Chem. Scripta,* 1972, **2**, 117-120.
17. Laemmli, U. K. *Nature,* 1970, **227**, 680-685.
18. Waddell, W. J. *J. Lab. Clin. Med.,* 1956, **48**, 311-314.

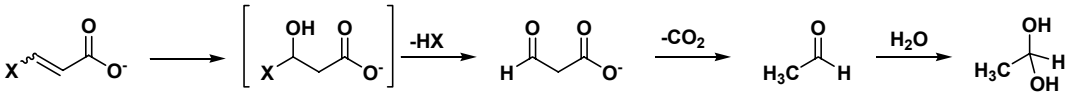
19. Whitman, C. P., Aird, B. A., Gillespie, W. R., and Stolowich, N. J., *JACS,* 1991, **113**, 3154–3162.
20. Johnson, W. H., Jr., Czerwinski, R. M., Stamps, S. L., and Whitman, C. P. *Biochemistry,* 1999, **38**, 16024-16033.
21. Person, M. D., Monks, T. J., and Lau, S. S. *Chem. Res. Toxicol*., 2003, in press.
22. O'Brien, P. J., and Herschlag, D. *Chem. Biol.,* 1999, **6**, R91-R105.
23. Braddon, S. A., and Dence, C. W. *Tappi,* 1968, **51**, 249-256.
24. Jencks, W. P. *Catalysis in Chemistry and Enzymology*, 1987, pp 116-120.

**Scheme 24**

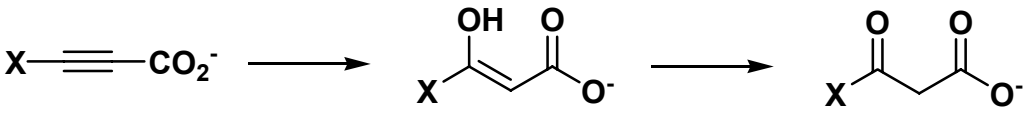



**1** **2** **3**

**Scheme 25**



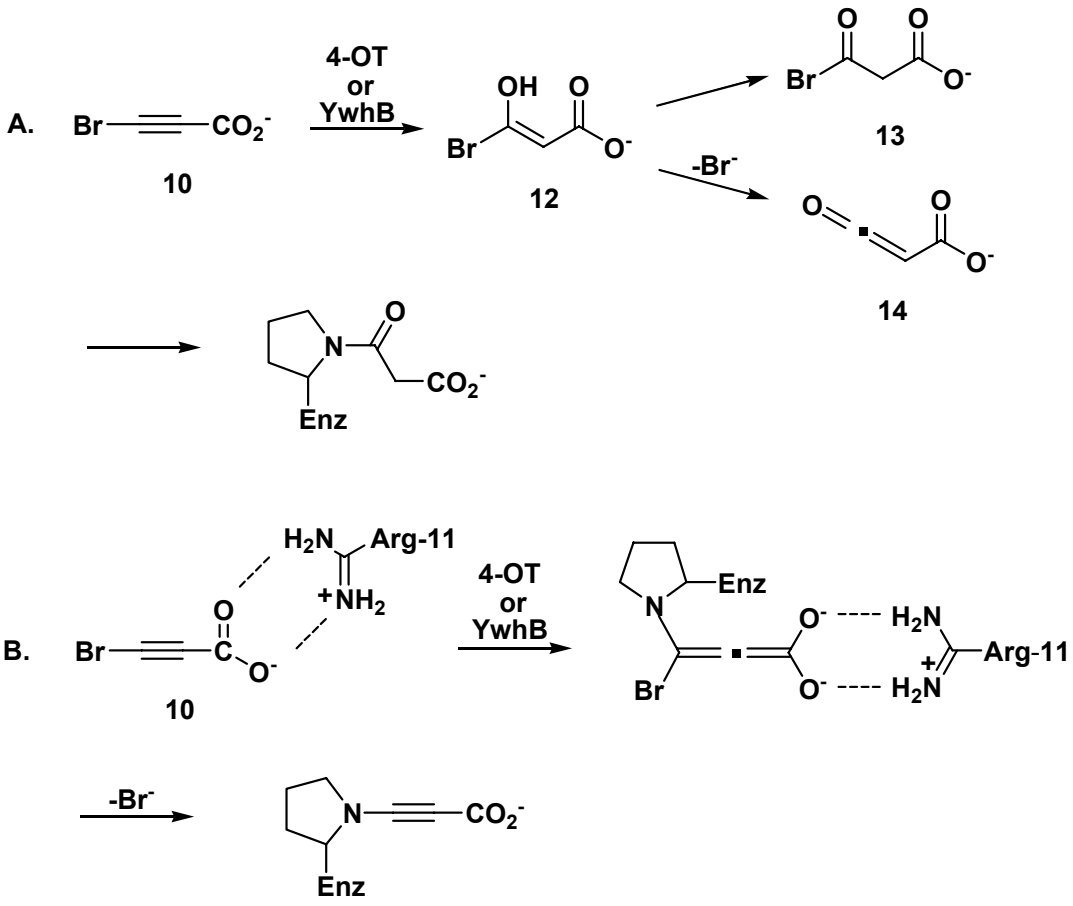
**Scheme 26**




**10: X = Br**

**11: X = Cl**

**Scheme 27**


A.
Br
CO2-
10
4-OT
or
YwhB
OH
O
Br
O-
12
O
O
Br
O-
13
-Br-
O
O
O-
14
N
O
CO2-
Enz
B.
H2N
Arg-11
+NH2
O
Br
C
O-
10
4-OT
or
YwhB
Enz
N
O-
H2N
Arg-11
Br
O-
H2N
-Br-
N
CO2-
Enz

**Scheme 28**

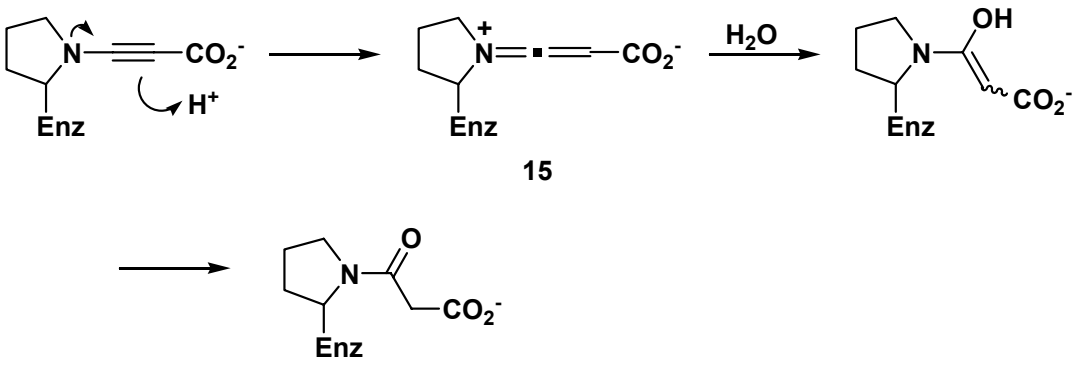
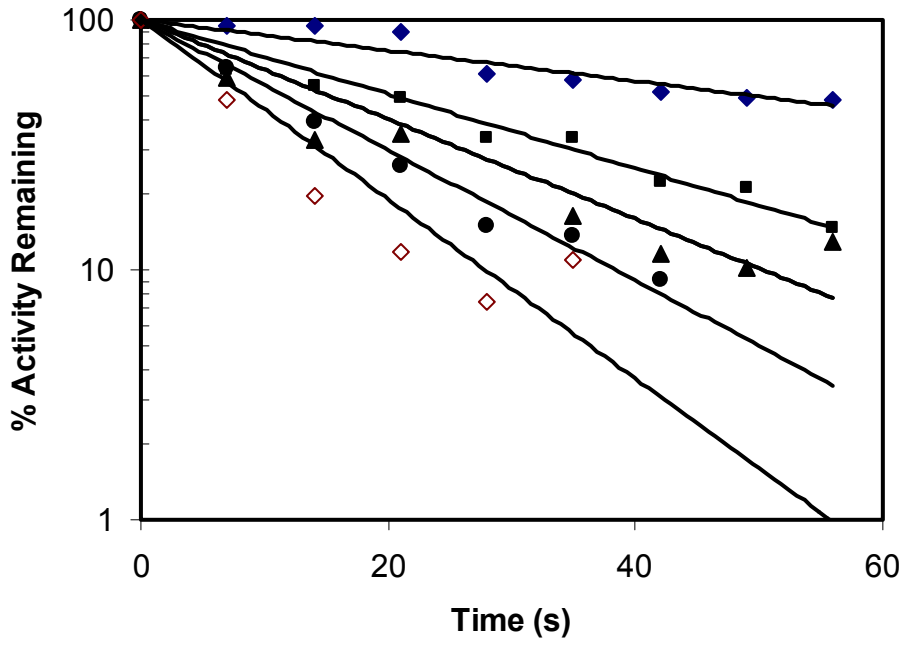


**Figure 37A.** The inactivation of 4-OT after incubation with varying amounts of **10** (filled diamonds, 80 μM; filled squares, 175 μM; filled triangles, 225 μM; filled circles, 400 μM; empty diamonds, 500 μM).

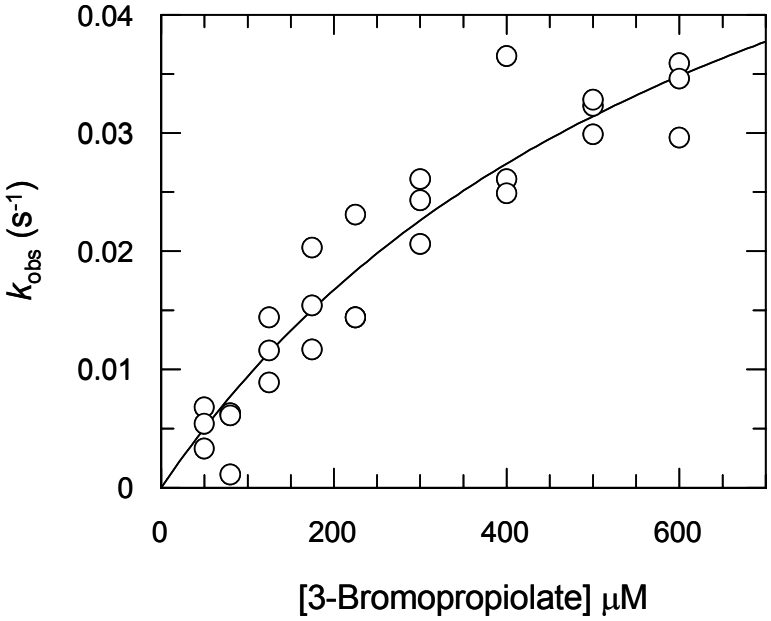
**Figure 37B**. Determination of the $k_{inact}$ and $K_I$ values for **10**. A plot of the $k_{obsd}$ values for inactivation measured in 27 experiments as a function of varying amounts of **10** (50-600 μM). The values of $k_{inact}$ and $K_I$ obtained from this plot are 0.08 ± 0.02 $s^{-1}$ and 706 ± 227 μM, respectively.

**Figure 37C.** Protection of 4-OT from inactivation by **10** using **1-3** (filled diamonds, 0 mM; filled squares, 0.41 mM; filled triangles, 1.04 mM).

**A**



**B**


0.04
0.03
0.02
0.01
0
$k_{obs}$ ($s^{-1}$)
0
200
400
600
[3-Bromopropiolate] μM

**C**

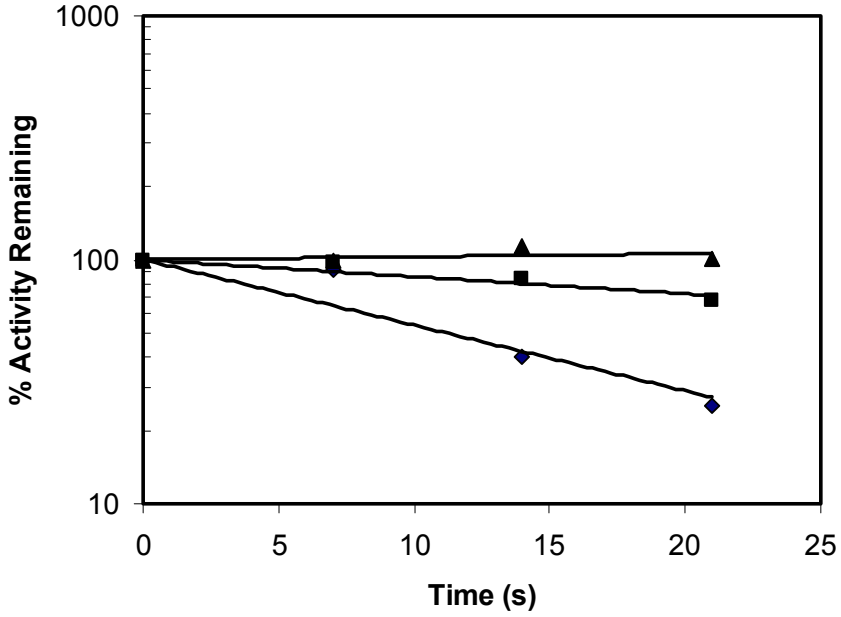
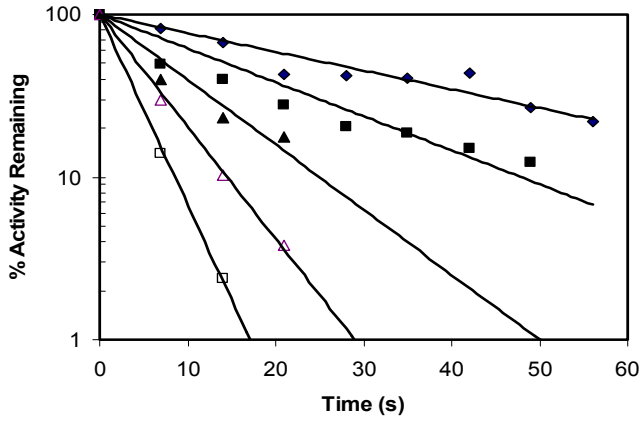

1000
100
10
% Activity Remaining
0
5
10
15
20
25
Time (s)

**Figure 38A.** The inactivation of 4-OT after incubation with varying amounts of **11** (filled diamonds, 50 μM; filled squares, 80 μM; filled triangles, 200 μM; empty triangles, 300 μM; empty squares, 500 μM).

**Figure 38B**. Determination of the $k_{inact}$ and $K_I$ values for **11**. A plot of the $k_{obsd}$ values for inactivation measured in 21 experiments as a function of varying amounts of **11** (50-750 μM). The values of $k_{inact}$ and $K_I$ obtained from this plot are 0.47 ± 0.16 $s^{-1}$ and 1.91 ± 0.85 mM, respectively.

**A**




**B**

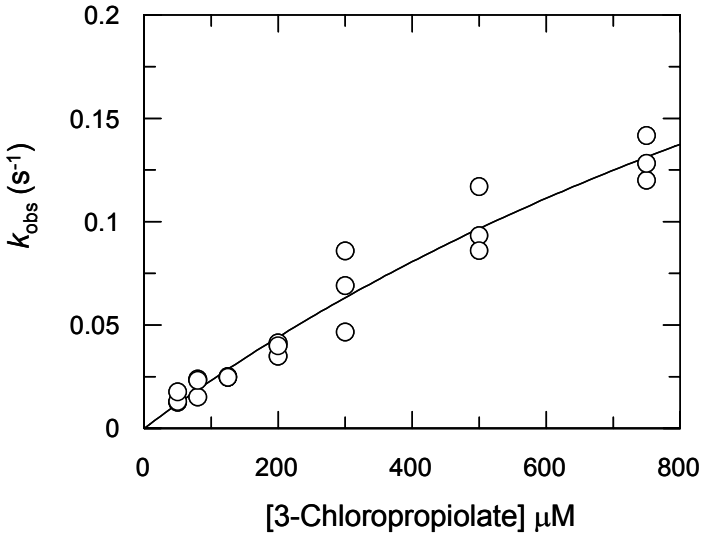
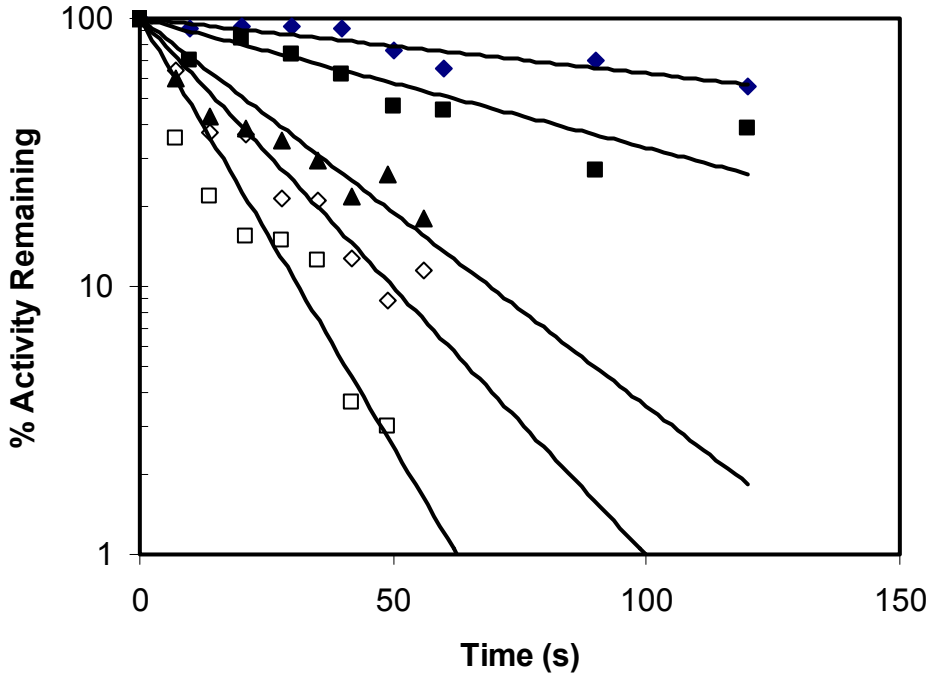


**Figure 39A.** The inactivation of YwhB after incubation with varying amounts of **10** (filled diamonds, 50 μM; filled squares, 100 μM; filled triangles, 300 μM; empty diamonds, 550 μM; empty squares, 1000 μM).

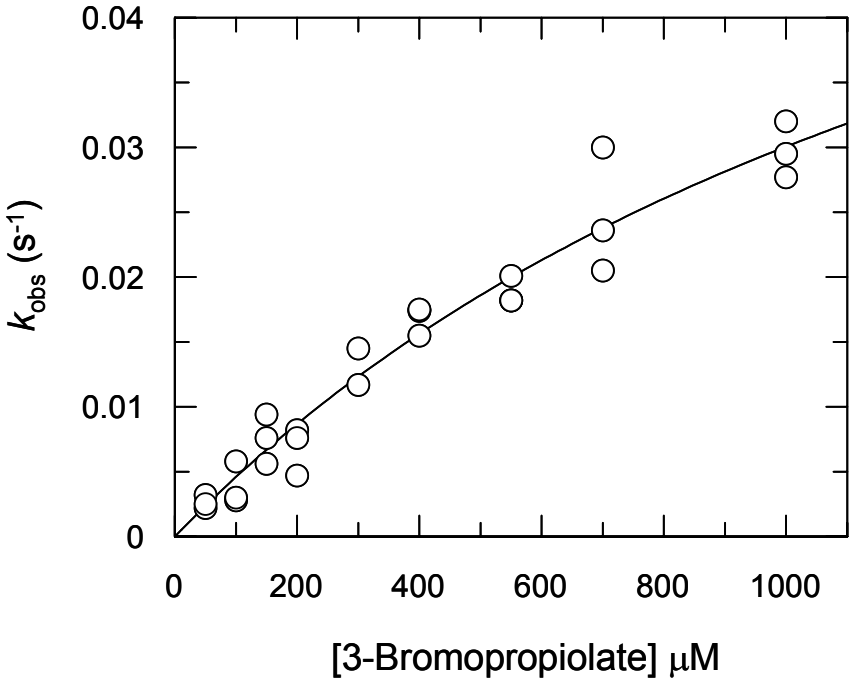
**Figure 39B**. Determination of the $k_{inact}$ and $K_I$ values for **10**. A plot of the $k_{obsd}$ values for inactivation measured in 27 experiments as a function of varying amounts of **11** (50-1000 μM). The values of $k_{inact}$ and $K_I$ obtained from this plot are 0.08 ± 0.02 $s^{-1}$ and 1.6 ± 0.5 mM, respectively.

**Figure 39C.** Protection of YwhB from inactivation by **10** using **1-3** (filled diamonds, 0 mM; filled squares, 0.835 mM; empty trianles, 1.67 mM, empty squares, 2.5 mM).

**A**



**B**


0.04
0.03
0.02
0.01
0
$k_{obs}$ ($s^{-1}$)
0
200
400
600
800
1000
[3-Bromopropiolate] μM

**C**

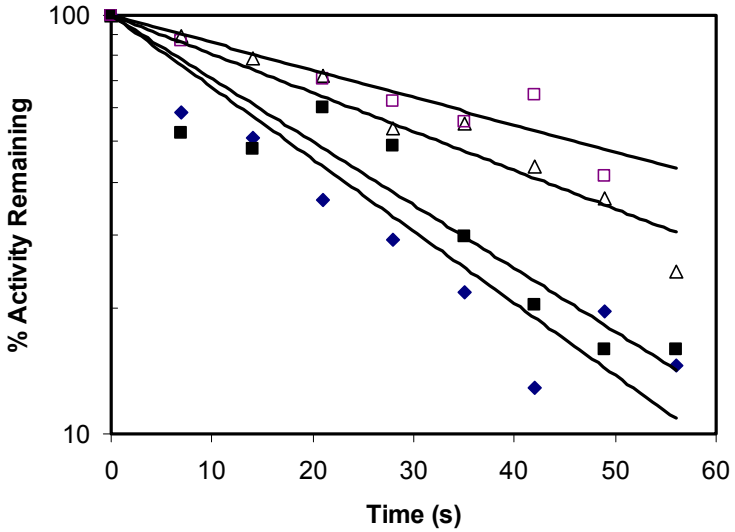
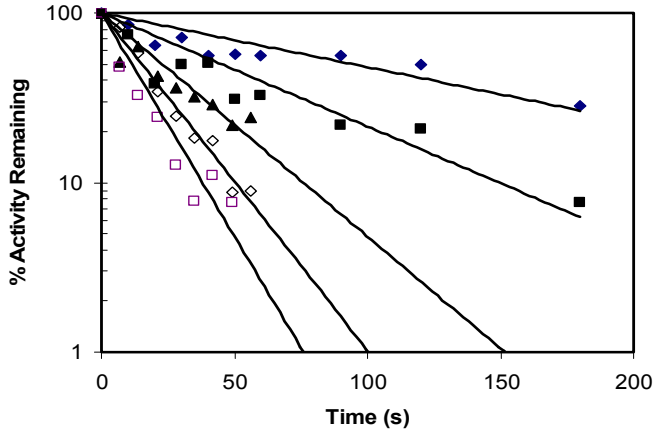

100
10
% Activity Remaining
0
10
20
30
40
50
60
Time (s)

**Figure 40A.** The inactivation of YwhB after incubation with varying amounts of **11** (filled diamonds, 50 μM; filled squares, 75 μM; filled triangles, 125 μM; empty diamonds, 175 μM; empty squares, 300 μM).

**Figure 40B**. Determination of the $k_{inact}$ and $K_I$ values for **11**. A plot of the $k_{obsd}$ values for inactivation measured in 36 experiments as a function of varying amounts of **11** (50-750 μM). The values of $k_{inact}$ and $K_I$ obtained from this plot are 0.11 ± 0.02 $s^{-1}$ and 875 ± 175 μM, respectively.

**A**




**B**

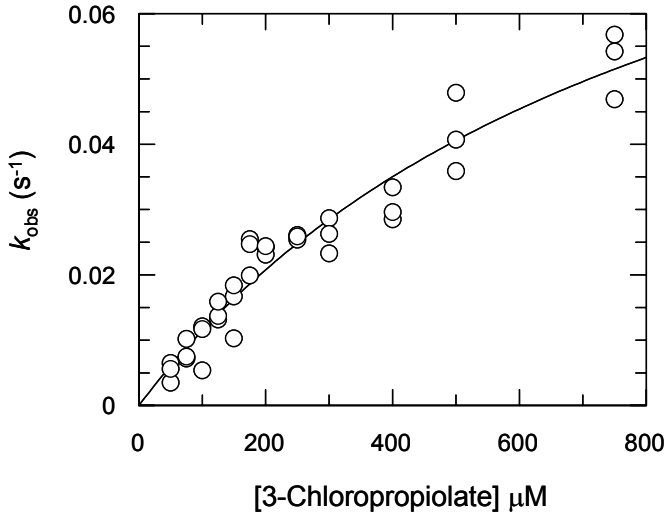


**Figure 41**. MALDI-PSD fragmentation spectra of the peak at 748.38 Da in the treated digest of YwhB, corresponding to the PYVTVK (1-6) fragment. The b ions are labeled with open triangles. The $b_1$ ions are shifted by +42 Da.

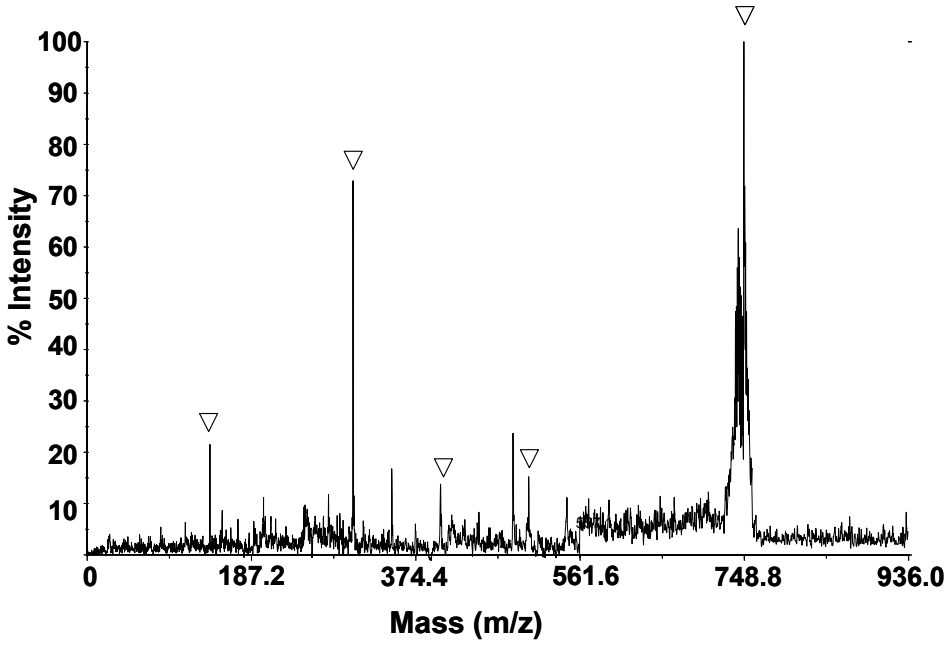


**Table 11. Dehalogenase activity of wild-type, synthetic, and mutant 4-OT enzymes.**

| Enzyme | Quantity (mg) | Substrate | Time (h) | % Conversion |
|---|---|---|---|---|
| Wild type | 0.64 | (*E*)-**5** | 136 | 74 |
| Wild type | 0.64 | (*E*)-**5** | 96 | 71 |
| Wild type | 0.19 | (*E*)-**5** | 139 | 64 |
| Synthetic | 0.36 | (*E*)-**5** | 208 | 67 |
| Synthetic | 0.20 | (*E*)-**5** | 208 | 22 |
| Wild type | 0.43 | (*E*)-**4** | 209 | 92 |
| R11A | 0.67 | (*E*)-**5** | 114 | 74 |
| R11A | 0.21 | (*E*)-**5** | 208 | 49 |
| P1A | 0.74 | (*E*)-**5** | 114 | 2 |
| Wild type | 0.64 | (*Z*)-**5** | 139 | 3 |
| Wild type | 0.43 | (*Z*)-**5** | 209 | 4 |

**Table 12. Dehalogenase activity of wild-type and mutant YwhB enzymes.**

| Enzyme | Quantity (mg) | Substrate | Time (h) | % Conversion |
|---|---|---|---|---|
| Wild type | 0.33 | (*E*)-**5** | 171 | 30 |
| Wild type | 0.33 | (*E*)-**4** | 616 | 93 |
| R11A | 0.60 | (*E*)-**5** | 137 | 5 |
| P1A | 0.54 | (*E*)-**5** | 108 | ND[a] |
| Wild type | 0.33 | (*Z*)-**5** | 137 | ND[a] |
| Wild type | 0.51 | (*Z*)-**5** | 209 | ND[a] |

[a]ND: Not detectable

**Table 13. Mass Spectral Analysis of V8 Digest of 4-OT.**

| *Peptide fragment* | *Calculated mass*[a] | *Observed mass (control)* | *Observed mass (treated)* |
|---|---|---|---|
| 1-9 | 1034.24 | 1033.59, 1055.69[b], 1071.64[b] | 1077.73, 1097.62[b], 1113. 68[b] |
| 10-17 | 948.96 | 948.44 | 948.47 |
| 18-22 | 631.75 | 631.38 | 631.40 |
| 23-44 | 2357.70 | 2356.22 | 2356.33 |
| 45-62 | 1915.27 | 1913.98 | 1914.07 |

[a]These values were calculated using the average molecular mass (Da).

[b]The +22 and +38 peaks are sodium and potassium adducts, respectively.

**Table 14. Mass Spectral Analysis of the V8 Digest of YwhB.**

| *Peptide fragment* | *Calculated mass*[a] | *Observed mass (control)* | *Observed mass (treated)* |
|---|---|---|---|
| 1-9 | 1080.33 | 1079.57 | ( )+ 42 (weak) |
| 1-25 | 2964.40 | 2962.49 | 3004.58 |
| 10-25 | 1903.10 | 1901.95 | 1901.98 |
| 30-44 | 1652.84 | 1651.80 (weak) | 1651.84 |
| 36-44 | 1105.29 | 1105.60 | 1105.61 |
| 36-61 | 3095.60 | 3093.59 | 3093.51 |
| 37-43 | 976.10 | 976.58 | 976.57 |
| 44-61 | 2136.43 | 2135.98 | 2136.02 |
| 45-61 | 2008.32 | 2006.95 | 2006.97 |

[a]These values were calculated using the average molecular mass (Da).

**BIBLIOGRAPHY**


1. Ahvazi, B., *et al. Biochem. J.*, 2000, **349**, 853-861.
2. Almrud, J. J., *et al. Biochemistry*, 2002, **41**, 12010-12024.
3. Anandarajah, K., Kiefer, P. M., Jr., Donohoe, B. S., and Copley, S. D. *Biochemistry*, 2000, **39**, 5303-5311.
4. Andersson, K. *Chem. Scripta,* 1972, **2**, 117-120.
5. Armstrong, R. N. *Chem.. Res. Toxicol.*, 1997, **10**, 2-18.
6. Babbitt, P. C. and Gerlt, J. A. *J. Biol. Chem.*, 1997, **272**, 30591-30594.
7. Babbitt, P. C., and Gerlt, J. A. *Adv. Prot. Chem.*, 2001, **55**, 1-28.
8. Babbitt, P. C., *et al. Biochemistry*, 1996, **35**, 16489-16501.
9. Babbitt, P. C., *et al. Science*, 1995, **267**, 1159-1161.
10. Bahnson, B. J. and Anderson, V. E. *Biochemistry*, 1989, **28**, 4173-4181.
11. Banner, D.W., *et al. Nature,* 1975, **255**, 609-14.
12. Bernhagen, J., Calandra, T., and Bucala, R., 1998, *J. Mol. Med.,* **76**, 151-61.
13. Blanco, J., Moore, R. A., Kabaleeswaran, V., and Viola, R. E., *Protein Science*, 2003, **12**, 27-33.
14. Board, P. G., Baker, R. T., Chelvanayagam, G., and Jermiin, L. S. *Biochem. J.*, 1997, **328**, 929-935.
15. Bognar, A., and Meighen, E. A., *J. Biol. Chem.*, 1978, **253**, 446-450.
16. Braddon, S. A., and Dence, C. W. *Tappi,* 1968, **51**, 249-256.
17. Busch, K. B., and Fromm, H. *Plant Phys.*, 1999, **121**, 589-597.
18. Cenedella, R. J. *Ophthalmic Res.*, 2001, **33**, 210-216.
19. Cerdan, P., Wasserfallen, A., Rekik, M., Timmis, K.N., and Harayama, S. *J. Bact.*, 1994, **176**, 6074-6081
20. Charter, J. H., II, *et al*. *Biochemistry,* 1974, **13**, 1227-1233.
21. Chen, L. H., *et al. J. Biol. Chem.*, 1992, **267**, 17716-17721.
22. Chirpich, T. P., Zappia, V., Costilow, R. N., and Barker, H.A. *J. Biol. Chem.*, 1970, **245**, 1778-1789.

23. Cobessi, D., *et al.*, *J. Mol Biol.*, 1999, **290**, 161-173.

24. Copley, S. D. *Trends. Biochem. Sci.*, 2000, **25**, 261-265.

25. Copley, S. D., in *Comprehensive Natural Products Chemistry,* Vol. 5, pp 401-422, 1999.

26. Costilow, R. N., Rochovansky, O. M., and Barker, H. A. *J. Biol. Chem.,* 1966, **241**, 1573-1578.

27. Czerwinski, R. M., *et al. Biochemistry,* 1997, **36**, 14551-14560.

28. Czerwinski, R. M., *et al. Biochemistry*, 1999, **38**, 12358-12366.

29. Czerwinski, R. M., Harris, T. K., Massiah, M. A., Mildvan, A. S., and Whitman, C. P., *Biochemistry*, 2001, **40**, 1984-1995.

30. D'Ordine, R. L., Bahnson, B. J., Tonge, P. J., and Anderson, V. E. *Biochemistry*, 1994, **33**, 14733-14742.

31. Duin, E. C., *et al*. *Biochemistry*, 1997, **36**, 11811-11820.

32. Fani, R., *et al. Gene,* 1997, **197,** 9-17.

33. Fani, R., Lio, P., and Lazcano, A. *J. Mol. Evol.*, 1995, **41**, 760-774.

34. Fani, R., Lio, P., Chiarelli, I., Bazzicalupo, M. *J. Mol. Evol.* 1994, 38: 489.

35. Fee., J. A., Hegeman, G. D., and Kenyon, G. L. *Biochemistry*, 1974, **13**, 2528-2532.

36. Fitzgerald, M. C., Chernushevich, I., Standing, K. G., Kent, S. B. H., and Whitman, C. P., *JACS*, 1995, **117**, 11075-11080.

37. Fitzgerald, M. C., Chernushevich, I., Standing, K. G., Whitman, C. P., and Kent, S. B. H., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 6851-6856.

38. Fong, W., and Choi, K., *Chem. Biol. Interact.*, 2001, **130-132**, 161-171.

39. Frey, P. A. and Moss, M. L. *Cold Spring Harbor Symp. Quant. Biol.*, 1987, **52**, 571–577.

40. Gaal, A. and Neujahr, H. Y. *J. Bact.*, 1979, **137**, 13-21.

41. Gerlt, J. A. and Babbitt, P. C. *Annu. Rev. Biochem.,* 2001, **70**, 209-246.

42. Gerlt, J. A., and Babbitt, P. C. *Curr. Opin. Chem. Biol.*, 1998, **2**, 607-612.

43. Gil, A. M., *et al. J. Agric. Food Chem.,* 2000, **48**, 1524-1536.

44. Gorini, L., Gundersen, W., and Burger, M. *Cold Spring Harbor Symp. Quant. Biol.,* 1961, **26**, 173.

45. Gribble, G. W. *Environ. Sci. Technol.*, 1994, **28**, 311A-319A.

46. Guerrero, F. D., Jones, J. T., and Mullet, J. E. *Plant Mol. Biol.*, 1990, **15**, 11-26.

47. Guthrie, J. P. *JACS,* 1972, **94**, 7020-7024.

48. Hadfield, A. T., *et al. J. Mol Biol.*, 1999, **289**, 991-1002.

49. Harris, T. K., *et al. Biochemistry*, 1999, **38**, 12343-12357.

50. Hartmans, S., Jansen, M. W., van der Werf, M. J., and De Bont, J. A. M., *J. Gen. Microbiol.*, 1991, **137**, 2025-2032.

51. Hempel, J., *et al. Adv. Exp. Med Biol.*, 1999, **463**, 53-59.

52. Hermanowski-Vosatka, A., *et al. Biochemistry*, 1999, **38**, 12841-12849.

53. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. *Gene,* 1989, **77**, 51-59.

54. Hofstein, H. A., Feng, Y., Anderson, V. E., and Tonge, P. J. *Biochemistry*, 1999, **38**, 9508-9516.

55. Holden, H. M., Benning, M. M., Haller, T., and Gerlt, J. A. *Acc. Chem. Res.*, 2001, **34**, 145-157.

56. Holler, T. P., Foltin, S. K., Ye, Q.-Z., and Hupe, D. J. *Gene,* 1993, **136**, 323-328.

57. Horn, J. M., Harayama, S. and Timmis, K. N., *Mol. Microbiol*., 1991, **5**, 2459-2474.

58. Horowitz, N. H. *Proc. Natl. Acad. Sci. USA,* 1945, **31**, 153-157.

59. Horowitz, N. H., in Evolving Genes and Proteins. 1965, 15-23.

60. Hurley, T. D., Steinmetz, C. G., and Weiner, H. *Adv. Exp. Med. Biol.*, 1999, **463**, 15-25.

61. Inoue, J., Shaw, J. P., Rekik, M., and Harayama, S. *J Bacteriol.*, 1995, **177**, 1196-201.

62. Jacob, F. and Monod, J. *Cold Spring Harbor Symp. Quant. Biol.,* 1961, **26**, 193.

63. Jencks, W. P. Catalysis in Chemistry and Enzymology, 1987, pp 116-120.

64. Jensen, R. A. *Annu. Rev. Microbiol.*, 1976, **30**, 409-425.

65. Johansson, K., *et al. Prot. Sci.,* 1998, **7**, 2106-2117.

66. Johnson, W. H. Jr., Czerwinski, R. M., Fitzgerald, M. C., and Whitman, C. P. *Biochemistry*, 1997, **36**, 15724-15732.

67. Johnson, W. H., Jr., Czerwinski, R. M., Stamps, S. L., and Whitman, C. P. *Biochemistry,* 1999, **38**, 16024-16033.

68. Johnson, W. H., Jr., Hajipour, G., and Whitman, C. P. *JACS*, 1995, **117**, 8719-8726.\

69. Karsten, W. E., and Viola, R.E. *Biochim. Biophys. Acta*, 1991, **1077**, 209-219.

70. Kita, A., *et al. Structure Fold Des.*, 1999, **7**, 25-34.

71. Kobayashi, T., *et al. J. Biochem. (Tokyo)*, 1995, **117**, 614-622.

72. Kosower, E. M. Physical Organic Chemistry, 1968, p 81.

73. Kunst, F., *et al.*, *Nature*, 1997, **390**, 249-256.

74. Laemmli, U.K. *Nature*, 1970, **227**, 680-685.

75. Lamb, A. L. and Newcomer, M. E. *Biochemistry*, 1999, **38**, 6003-6011.

76. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. *Science*, 2000, **289**, 1546-1550.

77. Lazcano, A. and Miller, S. L. *J. Mol. Evol.,* 1999, **49**, 424-431.

78. Lindahl, R. *Crit. Rev. Biochem. Mol. Biol.*, 1992, **27**, 283-335.

79. Liu, F., Cui, X., Horner, H. T., Weiner, H., and Schnable, P. S. *Plant Cell*, 2001, **13**, 1063-1078.

80. Liu, Z.-J., *et al. Nature Struc. Biol.*, 1997, **4**, 317-326.

81. Lundblad, R. L. and Noyes, C. M. Chemical Reagents for Protein Modification, 1984.

82. Luo, M. J., Smeland, T. E., Shoukry, K., and Schulz, H. *J. Biol. Chem.,* 1994, **269**, 2384-2388.

83. Maas, W. K. *Cold Spring Harbor Symp. Quant. Biol.,* 1961, **26**, 183.

84. Magnusson, O.T., Reed, G. H., and Frey, P.A. *Biochemistry*, 2001, **40**, 7773-7782.

85. Marcotte, P., and Walsh, C.T. *Biochemistry*, 1978, **17**, 5613-5619.

86. Marletta, M. A., Cheung, Y.-F., and Walsh, C. T. *Biochemistry,* 1982, **21**, 2637-2644.

87. McCarthy, D. L., Navarrete, S., Willett, W. S., Babbitt, P. C., and Copley, S. D. *Biochemistry*, 1996, **35**, 14634-14642.

88. Means, G.E., and Feeney, R.E. Chemical Modification of Proteins, 1971.

89. Miller, J. R., *et al. Biochemistry,* 2000, **39**, 15166–15178.

90. Murzin, A. G. *Curr. Opin. Struct. Biol.*, **6**, 386-394.

91. Nakai, C., *et al. J. Biol. Chem.*, 1983, **258**, 2923-2928.

92. Neidhardt, F. C., Block, P. L., and Smith, D. F. *J. Bact.*, 1974, **119**, 736-747.

93. Neidhart, D. J., Kenyon, G. L., Gerlt, J. A., and Petsko, G. A. *Nature*, 1990, **347**, 692-693.

94. Newman, J., *et al. Biochemistry,* 1999, **38**, 16105-16114.

95. O'Brien, P. J. and Herschlag, D. *Chem. Biol.*, 1999, **6**, R91-R105.

96. Ollagnier-de Choudens, S., *et al. Biochemistry,* 2000, **39**, 4165–4173.

97. Ou, L.-T., Thomas, J. E., Chung, K.-Y., and Ogram, A. V., *Biodegradation*, 2001, **12**, 39-47.

98. Palmer, D. R. J., *et al. Biochemistry,* 1999, **38**, 4252-4258.

99. Pereira, F., *et al. Biochem. Biophy. Res. Comm.*, 1991, **175**, 831-838.

100. Person, M. D., Monks, T. J., and Lau, S. S. *Chem. Res. Toxicol*., 2003 (in press).

101. Petsko, G. A., Kenyon, G. L., Gerlt, J. A., Ringe, D., and Kozarich, J. W. *Trends Biochem. Sci.*, 1993, **18**, 372-376.

102. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. *Science,* 2001, **291**, 1790-1792.

103. Pietruszko, R., Kikonyogo, A., Chern, M.K., and Izaguirre, G. *Adv. Exp. Med. Biol.,* 1997, **414**, 243-252.

104. Poelarends, G. J., Kulakov, L. A., Larkin, M. J., van Hylckama Vlieg, J. E. T., and Janssen, D. B. *J. Bacteriol.,* 2000, **182**, 2191-2199.

105. Poelarends, G. J., Saunier, R., and Janssen, D. B., *J. Bact.*, 2001, **183**, 4269-4277.

106. Poelarends, G. J., Wilkens, M., Larkin, M. J., van Elsas, J. D., and Janssen, D. B., *Appl. Environ. Microbiol.*, 1998, **64**, 2931-2936.

107. Reed, L. J., and Hackert, M. L. *J. Biol. Chem.*, 1990, **265**, 8971-8974.

108. Roper, D. I., Fawcett, T., and Cooper, R. A., *Mol. Gen. Genet.*, 1993, **23**, 241-250.

109. Rosengren, E., *et al. Mol. Med.*, 1996, **2**, 143-149.

110. Rosengren, E., *et al. FEBS Lett.,*1997,**417**, 85-88.

111. Rossmann, M. G., Moras, D., and Olsen, K. W. *Nature*, 1974, **250**, 194-199.

112. Sambrook, J., Fritsch, E.F., and Maniatis, T. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, 1989.

113. Satya Narayan, V., and Nair, P. M. *Arch. Biochem. Biophys.,* 1989, **275**, 469-477.

114. Shaw, J. P., and Harayama, S. *Eur. J. Biochem.*, 1990, **191**, 705-714.

115. Shaw, J. P., Schwager, F., and Harayama, S. *Biochem. J.*, 1992, **283**, 789-794.

116. Shine, N. R., *et al. Bioorganic Chem.,* 2002, **30**, 249-263.

117. Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F., and Miller, N. E. *Nucl. Acids. Res.* 2001, **29,** 1097-1106.

118. Sparnins, F. L., Chapman, P. J., and Dagley, S., *J. Bact.*, 1974, **120**, 159-167.

119. Stadtman, T. C. *AdV. Enzymol. Relat. Areas Mol. Biol.*, 1973, **38**, 413-448.

120. Stamps, S. L., Fitzgerald, M. C., and Whitman, C. P. *Biochemistry,* 1998, **37**, 10195-10202.

121. Stanley, T. M. *et al*., *Biochemistry,* 2000, **39**, 718-726.

122. Stivers, J. T., Abeygunawardana, C., Mildvan, A. S., Hajipour, G., and Whitman, C.P. *Biochemistry,* 1996, **35**, 814-823.

123. Stivers, J. T., *et al*. *Biochemistry*, 1996, **35**, 803-813.

124. Strauss, F., Kollek, L., and Heyn, W. *Chem. Ber,* 1930, **63**, 1868-1899.

125. Stroeher, V. L., Boothe, J. G., and Good, A. G. *Plant Mol. Biol.*, 1995, **27**, 541-551.

126. Subramanya, H. S., *et al. Biochemistry*, 1996, **35**, 792-802.

127. Taylor, A. B., Czerwinski, R. M., Johnson, W. H. Jr., Whitman, C. P., and Hackert, M. L. *Biochemistry*, 1998, **37**, 14692-14700.

128. Taylor, K. L., *et al. Biochemistry*, 1995, **34**, 13881-13888.

129. Taylor, K. L., Xiang, H., Liu, R. Q., Yang, G., and Dunaway-Mariano, D. *Biochemistry*, 1997, **36**, 1349-1361.

130. Teunissen, P. J., Swarts, H. J., and Field, J. A. *Appl. Microbiol. Biotechnol.*, 1997, **47**, 695-700.

131. Thoma, R., Schwander, M., Liebl, W., Kirschner, K., Sterner, R. *Extremophiles,* 1998, **2,** 379.

132. Ting, H. H., and Crabbe, M. J., *Biochem. J.*, 1983, **215**, 361-368.

133. Van Hylckama Vlieg, J. E. T. and Janssen, D. B., *Biodegradation*, 1992, **2**, 139-150.

134. Vasiliou, V., and Sophos, N. A. *Chem. Biol. Interact.*, 2003, **143-144**, 5-22.

135. Vasiliou, V., Pappa, A., and Petersen, D. R. *Chem. Biol. Interact.*, 2000, **129**, 1-19.

136. Verschueren, K. H. G., Seljée, F., Rozeboom, H. J., Kalk, K. H., and Dijsktra, B. W. *Nature,* 1993, **363**, 693-698.

137. Vogel, H. J. *Cold Spring Harbor Symp. Quant. Biol.,* 1961, **26**, 163.

138. Waddell, W.J. *J. Lab. Clin. Med.,* 1956, **48**, 311-314.

139. Wang, S. C., Person, M. D., Johnson, W. H., Jr., and Whitman, C. P. *Biochemistry*, 2003, **42**, 8762-8773.

140. Wang, X., and Weiner, H., *Biochemistry*, 1997, **34**, 237-243.

141. Whitman, C. P., Aird, B. A., Gillespie, W. R., and Stolowich, N. J., *JACS,* 1991, **113**, 3154–3162.

142. Whitman, C. P., *Arch. Biochem. Biophys.*, 2002, **402**, 1-13.

143. Whitman, C. P., *et al. JACS,* 1992, **114**, 10104-10110.

144. Wu, W., *et al. Biochemistry*, 2000, **39**, 9561–9570.

145. Xiang, H., Dong, J., Carey, P. R., and Dunaway-Mariano, D. *Biochemistry*, 1999, **38**, 4207-4123.

146. Xiang, H., Luo, L. S., Taylor, K. L., and Dunaway-Mariano, D. *Biochemistry*, 1996, **35**, 8103-8109.

147. Xiang, H., Luo, L., Taylor, K. L., and Dunaway-Mariano, D. *Biochemistry*, 1999, **38**, 7638-7652.

148. Xu, L., Resing, K., Lawson, S. L., Babbitt, P. C., and Copley, S. D. *Biochemistry*, 1999, **38**, 7659-7669.

149. Xun, L., and Orser, C. S. *J. Bact.*, 1991, **173**, 4447-4453.

150. Xun, L., Topp, E., and Orser, C. S. *J. Bact.*, 1992, **174**, 8003-8007.

151. Yamauchi, K., Nakajima, J., Hayashi, H., Horiuchi, R., and Tata, J. R. *J. Biol. Chem.*, 1999, **274**, 8460-8469.

152. Yang, G., *et al*. *Biochemistry*, 1996, **35**, 10879-10885.

153. Yang, G., Liang, P. H., Dunaway-Mariano, D. *Biochemistry,* 1994, **33**, 8527-8531.

154. Ycas, M. *J. Theor. Biol.*, 1974, **44**, 145-160.

155. Ye, Q.-Z., Johnson, L. L., and Baragi, V. *Biochem. Biophys. Res.Commun.,* 1992, **186**, 143-149.

156. Zhang, L, Ahvazi, B., Szittner, R., Vrielink, A., and Meighen, E. *Biochemistry,* 2000, **39**, 14409-14418.

## VITA

Susan C Wang was born on April 14, 1976 in Manhasset, New York, the daughter of Lin-Lan Chiu and Chi-Rong Wang. At the age of three, she moved to the Cleveland, Ohio area, which she considers her home. She attended North Olmsted High School, graduating in 1994 as valedictorian. She attended Case Western Reserve University from 1994-1998. She graduated *summa cum laude* with a Bachelor's of Science degree in May 1998, majoring in biology and minoring in chemistry, computer science, and English. In August 1998, she enrolled in the Division of Medicinal Chemistry of the College of Pharmacy at The University of Texas at Austin. She was a doctoral student from 1998-2003 under the supervision of Christian P. Whitman, graduating in August 2003 with a Doctor of Philosophy in Pharmacy (Medicinal Chemistry).

## PUBLICATIONS


1. Stamps, S. L., Taylor, A. B., Wang, S. C., Hackert, M. L., and Whitman, C. P., "Mechanism of the Phenylpyruvate Tautomerase Activity of Macrophage Migration Inhibitory Factor: Properties of the P1G, P1A, Y95F, and N97A Mutants," *Biochemistry*, *39*, 9671 (2000).
2. Almrud, J. J., Kern, A. D., Wang, S. C., Czerwinski, R. M., Johnson, W. H., Jr., Murzin, A. G., Hackert, M. L., and Whitman, C. P. "The Crystal Structure of YdcE, a 4-Oxalocrotonate Tautomerase Homologue from *Escherichia coli*,

Confirms the Structural Basis for Oligomer Diversity," *Biochemistry, 41*, 12010 (2002).

3. Shine, N. R., Wang, S. C., Konopka, K., Burks, E. A., Duzgunes, N., and Whitman, C. P. "Secretory Leukocyte Protease Inhibitor: Inhibition of Human Immunodeficiency Virus-1 Infection of Monocytic THP-1 Cells by a Newly Cloned Protein," *Bioorganic Chemistry*, *30*, 249 (2002).
4. Wang, S. C., Person, M. D., Johnson, W. H., Jr., and Whitman, C. P. "Reactions of *trans*-3-Chloroacrylic Acid Dehalogenase with Acetylene Substrates: Consequences of and Evidence for a Hydration Reaction," *Biochemistry*, *42*, 8762 (2003).
5. Wang, S. C., Johnson, W. H., Jr., and Whitman, C. P. "The 4-Oxalocrotonate Tautomerase- and YwhB-catalyzed Hydration of 3*E*-Haloacrylates: Implications for the Evolution of New Enzymatic Activities," *Journal of the American Chemical Society*, submitted.

Permanent address: 1322 Waldorf Boulevard, #105, Madison, Wisconsin 53719

This dissertation was typed by the author.